AD

Award Number: DAMD17-99-1-9390

TITLE: Statistical Analysis of Multivariate Interval-Censored Data in Breast Cancer Follow-Up Studies

PRINCIPAL INVESTIGATOR: George Y.C. Wong, Ph.D.

CONTRACTING ORGANIZATION: Strang Cancer Prevention Center
New York, New York 10021-4601

REPORT DATE: July 2003

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

20031212 151

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
July 2003

3. REPORT TYPE AND DATES COVERED
Final (1 Jul 99-30 Jun 03)

4. TITLE AND SUBTITLE
Statistical Analysis of Multivariate Interval-Censored Data in Breast Cancer Follow-Up Studies

5. FUNDING NUMBERS
DAMD17-99-1-9390

6. AUTHOR(S)
George Y.C. Wong, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Strang Cancer Prevention Center
New York, New York 10021-4601

E-Mail: gwong@strang.org

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

The overall objective of our research proposal is nonparametric inference of the joint survival function $S(x_1, ..., x_d) = \Pr(X_1 > x_1, ..., X_d > x_d)$ of d (≥ 2) correlated time-to-event variables X_1, ..., X_d, each of which is subject to interval censoring. The standard estimator of S is the generalized maximum likelihood estimator (GMLE) $\hat{S}$. However, $\hat{S}$ cannot be expressed in a closed-form expression and its statistical properties have not been studied in the multivariate case. The technical objectives of this pioneer methodological research proposal are to develop asymptotic generalized maximal likelihood (GML) inference of S and to derive efficient computational algorithms for the GML procedure. In our fourth and final year of research, we have implemented a computer software for asymptotic inference of GMLE $\hat{\rho}$ of the correlation coefficient ρ between a pair of the X' variables. When the censoring distribution is continuous, we have numerically established $\hat{\rho}$ is not asymptotically normal and we have implemented a bootstrap method for obtaining interval estimator of ρ. Thus in our four years of research, we have successfully completed the tasks we proposed. The results will be useful to breast cancer researchers pursuing chemoprevention intervention trials involving multiple surrogate endpoint biomarkers, and genetic epidemiologists conducting studies on familial aggregation of breast cancer and related cancers.

14. SUBJECT TERMS
Breast cancer, multivariate interval-censored data, generalized maximum likelihood, consistency, asymptotic normality and efficiency.

15. NUMBER OF PAGES
67

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinion, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

_X_ Where copyrighted material is quoted, permission has been obtained to use such material.

_X_ Where material from documents designated for limited distribution is quoted, permission has be obtained to use the material.

_X_ Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals", prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institute of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of re-research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

George Wong
7/30/03
Principal Investigator

A. TABLE OF CONTENTS

B. INTRODUCTION

Interval-censored (IC) data are encountered in three areas of breast cancer research. The most common application is in clinical relapse follow-up studies in which the study endpoint is disease-free survival. When a patient relapses, it is usually known that the relapse takes place between two follow-up visits, and the exact time to relapse is unknown. In statistics, we say relapse time is interval censored. Interval censoring is also encountered in breast cancer registry studies in which information on family history of cancer is updated periodically. The Strang Breast Surveillance Program for women at increased risk for breast cancer, for instance, has enlisted over 800 women with complete pedigree information which is verified and updated continuously. Family history data such as age at diagnosis of a specific cancer, or a benign but risk-conferring condition, are obtained from each registrant at each update. Time to a cancer event, and definitely time to first detection of a benign condition, are at best known to fall in the time interval between the last update and age at diagnosis. A third but increasingly important area of application of interval censoring is in breast cancer chemoprevention experiments or prevention trials, which involve the observation of one or more surrogate endpoint biomarkers (SEB) over time. The scientific question of interest here is the estimation of time for the SEB to reach a target value, and time from cessation of intake of a chemopreventive agent to the loss of its protective effect. Unfortunately, the exact values of both these time variables are known only to lie in between two successive assay inspection times.

Let X denote a time-to-event variable with distribution $F(x) = Pr(X \leq x)$, or equivalently, survival function $S(x) = 1 - F(x)$. In interval censoring, X is not observed and is known only to lie in an observable interval (L, R). In our previous DOD funded grant, we have made fundamental contributions to both the theory of the generalized maximum likelihood (GML) estimation of S, and the computation in connection with the inference of GML estimator (GMLE) $\hat{S}$ of S. These contributions are restricted to the case of univariate interval-censored data.

Multivariate interval censoring involves $d \geq 2$ correlated X variables, each of which is subject to interval censoring. The main statistical concern here is the GML estimation of the joint survival function $S(x_1, ..., x_d) = Pr(X_1 > x_1, ..., X_d > x_d)$, and the correlations among the variables. Our interest in multivariate IC data is driven by needs arising from two related areas of breast cancer research at Strang. First, our investigators in the Strang Cancer Genetics Program want to study various patterns of familial aggregation of breast, ovarian and other forms of cancer using family history data from the Strang Breast

Surveillance Program. Studies of familial early onset of breast cancer, breast-ovarian and breast-prostate associations will lead to multivariate IC data of high dimensions; therefore, a proper statistical procedure together with a feasible software to deal with such data are very much needed. Second, we are conducting a one-year chemoprevention trial of indole-3-carbinol (I3C) for breast cancer prevention. In this prevention trial we are monitoring the levels of two SEB's, a urinary estrogen metabolite ratio and a blood counterpart, both of which are subject to interval censoring. An earlier dose-ranging study of I3C conducted by Wong *et al* [1] has been published.

Statistical analysis of multivariate IC data has never been attempted. In the multivariate situation, modeling of the intercorrelated time-to-event variables and their dependency structure will require a great deal of innovative thinking; moreover, GML computation in realistic sample sizes can be prohibitively difficult.

The overall aim of this research proposal is to develop statistical inference for multivariate interval-censored data that are encountered in breast cancer chemoprevention trials employing multiple surrogate endpoint biomarkers, and in breast cancer registry follow-up studies of familial aggregation of breast and other forms of cancer. Asymptotic generalized maximum likelihood theory has been investigated and computer software package for maximum likelihood inference and Kaplan-Meier type survival plots has been implemented.

C. BODY

Consider nonparametric estimation of the joint survival function $S(x_1, ..., x_d) = \Pr(X_1 > x_1, ..., X_d > x_d)$ of $d \geq 2$ intercorrelated time-to-event variables X_1, ..., X_d, each of which is subject to interval censoring. For ease of presentation and without any loss of generality, we shall restrict our discussion to the bivariate case $\underline{X} = (X_1, X_2)$.

Let (U_i, V_i) denote two consecutive follow-up times corresponding to X_i, and (L_i, R_i) denote the <u>observable</u> interval-censored (IC) data for X_i defined as

$$(L_i, R_i) = \begin{cases} (0, U_i) & \text{if } X_i \leq U_i, \\ (U_i, V_i) & \text{if } U_i < X_i \leq V_i, \\ (V_i, +\infty) & \text{if } X_i > V_i, \end{cases} \tag{1}$$

for $i = 1$, 2. Under this two-dimensional interval censorship model, data are always interval censored, i.e., $L_i < R_i$ with probability one. If we allow the possibility of having exact observations in the data, so that

$$L_i = R_i = X_i, \tag{2}$$

then (1) and (2) together define a two-dimensional mixed interval censorship model.

Let B_i denote any one of $[0, U_i]$, $(U_i, V_i]$ and $(V_i, +\infty)$. Therefore, a bivariate IC data point is a rectangular region in $\mathcal{R}^2$ taking one of the nine forms in $\mathcal{B} = \{B_k \times B_l : k, l = 1, 2, 3\}$. Given a sample of size n, the observations $(L_{i1}, R_{i1}, L_{i2}, R_{i2})$ can be represented by rectangle subsets $I_i \in \mathcal{B}$, for $i = 1, ..., n$. Define a maximal intersection (MI) A of the observable rectangles $I_1, ..., I_n$, to be a nonempty finite intersection of the I_i's such that $A \cap I_i = \emptyset$ or A, for each i. Let $A_1, ..., A_m$, denote the distinct maximal intersections with respect to $I_1, ..., I_n$.

The generalized likelihood function of S is given by $\Lambda_n = \mu_S(I_1) \times \cdots \times \mu_S(I_n)$, where $\mu_S(\cdot)$ is the probability measure induced by S. Wong and Yu [2] show that the GMLE $\hat{S}$, which maximizes Λ_n, must assign all the probability masses $s_1, ..., s_m$ to $A_1, ..., A_m$. In general, $\hat{S}$ has to be obtained iteratively. Since $\hat{S}$ is also a self-consistent estimate (SCE), we can implement the SCE algorithm by solving for $\hat{s}_1, ..., \hat{s}_m$ in

$$s_j = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_{ij} s_j}{\sum_{k=1}^{m} \delta_{ik} s_k},$$

$j = 1, ..., m$, where $\delta_{ij} = \mathbf{1}[A_j \subset I_i]$, $\mathbf{1}[\cdot]$ denoting the indicator function, and obtain an SCE of $S(\underline{x})$

$$\tilde{S}(\underline{x}) = \sum_{A_j \subset (x_1, +\infty) \times \cdots \times (x_d, +\infty)} \hat{s}_j.$$

With starting values $s_j^{(0)} = 1/m$ for all j, $\tilde{S}(\underline{x})$ is the GMLE at convergence.

In the **first** and **second** years of our research, we established consistency of the GMLE $\hat{S}$ under both discrete and continuous assumptions. We also established asymptotic normality of the GMLE $\hat{S}$ under a set of discrete assumptions. Additionally, we derived asymptotic properties of the weighted Kaplan-Meier test statistics given by

$$D = \int_{\underline{x} \geq 0} W(\underline{x})(\hat{S}_A(\underline{x}) - \hat{S}_B(\underline{x})) d\underline{x},$$

where $W(\cdot)$ is a given weight function, and A and B refer to two comparison conditions.

When the underlying distribution F_0 and the distribution of the censoring variables are both continuous, Groeneboom and Wellner [5] have conjectured in the univariate case that $\hat{S}$ is not asymptotically normally distributed and the convergence rate of $\hat{S}$ is of order $(n\ln n)^{1/3}$. We expect the same observation to hold true in the multivariate situation. Because of the theoretical difficulty with establishing the asymptotic non-normal distribution of $\hat{S}$ under continuous distribution assumption, we have to resort to the bootstrap method numerically evaluate the asymptotic inference of $\hat{S}$.

We have devoted our effort to this aspect of research in the **fourth** year of our DOD grant. We have develop a computer program to perform the bootstrap asymptotic calculations. The program is made available to the public via the internet at www.math.binghamton.edu/qyu/index.html. We have also carried out simulation studies to investigate whether the bootstrap method can provide a consistent estimate of the standard deviation of $\hat{S}$ under uniform distributions for sample sizes 50, 100 and 200. Let SD_n denote the bootstrap estimate of the standard deviation with sample size n. Our simulation results suggest that (1) $\hat{S}$ converges in distribution at the rate of $(n\ln n)^{1/3}$ and (2) SD_n will be sufficiently close to the standard deviation when the sample size is at least 50.

In the **third** and **fourth** and final years of our research, we have studied the consistency property of $\hat{S}$ under more general conditions. A manuscript summarizing the findings have just been submitted to a statistical journal [3].

Also, in the **fourth** year of our research, we have updated and expanded a computer software package for carrying out asymptotic GML inference of $\hat{S}$. The package is made available for the public via the internet at www.math.binghamton.edu/qyu/index.html.

A key feature of multivariate IC data and a parameter of substantive importance is the correlation coefficient ρ between a pair of the X variables, say X_1 and X_2. The GMLE of $\rho(X_1, X_2)$ is
$\hat{\rho}(x_1, x_2)$

$$= \frac{\int\int x_1 x_2 d\hat{F}(x_1,x_2) - \int\int x_1 d\hat{F}(x_1,x_2)\int\int x_2 d\hat{F}(x_1,x_2)}{\{[\int\int x_1^2 d\hat{F}(x_1,x_2) - (\int\int x_1 d\hat{F}(x_1,x_2))^2][\int\int x_2^2 d\hat{F}(x_1,x_2)(\int\int x_2 d\hat{F}(x_1,x_2))^2]\}^{1/2}}.$$

In a follow-up study involving interval censoring, it is often the case that not all events will take place by the end of the study. In this situation, $\hat{\rho}$ will not provide a consistent estimate of ρ. Let τ denote the largest follow-up time. A more appropriate correlation coefficient to consider is

$$\rho_\tau(x_1, x_2) = \frac{Cov(X_1, X_2 | X_1, X_2 \le \tau)}{\sqrt{Var(X_1|X_1 \le \tau) Var(X_2|X_2 \le \tau)}}.$$

$\hat{F}$ $(= 1 - \hat{S})$, the GMLE of F_o $(= 1 - S)$, is a discrete cdf with discontinuity points at the upper-right vertexes of the maximum intersections. Without loss of generality, let $a_1 < \cdots < a_m$ be the set of partition points of the real line such that the set $\{(a_i, a_j) : i, j \in \{0, 1, ..., m, m+1\}\}$ contains all the discontinuity points of $\hat{F}$, where $a_0 = -\infty$ and $a_{m+1} =$

∞. Let $\hat{s}_{ij}$ denote the GMLE of the bivariate probability weight assigned to (a_i, a_j) by $\hat{F}$. The GMLE of ρ_τ is given by

$$\hat{\rho}_\tau = \frac{E_{00}E_{12} - E_{10}E_{02}}{\sqrt{[E_{00}E_{11} - (E_{10})^2][E_{00}E_{22} - (E_{02})^2]}},$$

where $E_{12} = \sum_{a_i,a_j<\infty} a_i a_j \hat{s}_{ij}$, $E_{00} = \sum_{a_i,a_j<\infty} \hat{s}_{ij}$, $E_{10} = \sum_{a_i,a_j<\infty} a_i \hat{s}_{ij}$, $E_{02} = \sum_{a_i,a_j<\infty} a_j \hat{s}_{ij}$, $E_{11} = \sum_{a_i,a_j<\infty} a_i^2 \hat{s}_{ij}$, and $E_{22} = \sum_{a_i,a_j<\infty} a_j^2 \hat{s}_{ij}$.

From the consistency results of Wong and Yu [2], and Yu, Yu and Wong [3] we can show that $\hat{\rho}_\tau$ is consistent under the assumption that the union of the support sets of censoring variables is dense. Moreover, if the range of the censoring vector is finite, $\hat{\rho}_\tau$ can be shown to be asymptotically normally distributed. The asymptotic variance of $\hat{\rho}_\tau$ can be estimated by

$$\hat{\sigma}^2 = B\mathcal{I}^{-1}B',$$

where $B = \frac{\partial \rho_\tau}{\partial \mathrm{s}}$, $\mathrm{s} = \{s_{ij} : \ (i,j) \neq (m,m)\}'$, and $\mathcal{I}$ is the information matrix, that is

$$\mathcal{I} = -\frac{\partial^2 \ln \mathrm{L}}{\partial \mathrm{s}' \partial \mathrm{s}}.$$

We are preparing a manuscript on the asymptotic properties of $\hat{\rho}_\tau$.

When the finite distribution assumption regarding the censoring vector is not met, the expression for $\hat{\sigma}^2$ given above is no longer a consistent estimator of the variance of the GMLE $\hat{\rho}_\tau$ of the correlation coefficient ρ_τ. As in the case of $\hat{S}$, we have devoted our effort in the **fourth** year of research to investigate the asymptotic behavior of $\hat{\rho}_\tau$ using the bootstrap method. Again, we have established that the asymptotic behavior of $\hat{\rho}_\tau$ is similar to that of $\hat{S}$. Our research suggests that the bootstrap method is an important practical statistical tool that can be easily used to obtain interval estimate of the correlation coefficient $\hat{\rho}_\tau$. We have made available the bootstrap computer program for $\hat{\rho}_\tau$ to the public via the internet at www.math.binghamton.edu/qyu/index.html.

D. KEY RESEARCH ACCOMPLISHMENTS

- We have implemented a computer software package for calculating the GMLE $\hat{S}$ of the joint survival function $S(x_1, ..., x_d) = Pr(X_1 > x_1, ..., X_d > x_d)$ of $(d \geq 2)$ correlated time-to-event variables $X_1, ..., X_d$, each of which is subject to interval censoring.

- We have established consistency of $\hat{S}$ under both discrete and continuous distributional assumptions. We have also investigated consistency of $\hat{S}$ under a range of conditions defined by weaker assumptions.

- We have established asymptotic normality for $\hat{S}$ under finite distributional assumptions, and pointed out $\hat{S}$ may not converge in distribution to a normal variable under continuous assumptions.

- We have also encountered and provided a solution to a methodological problem arising from an unexpected finding that $\hat{S}$ may not be unique in the case of multivariate interval censoring.

- We have established consistency for the GMLE $\hat{\rho}_\tau$ of the correlation coefficient ρ_τ between a pair of correlated time-to-event variables, both of which are subject to interval censoring. Under finite distributed assumptions, we have derived the asymptotic normality of $\hat{S}$.

- When finite distributional assumptions are inappropriate, we have implemented a bootstrap method to obtain interval estimate of $\hat{\rho}_\tau$. Through simulation studies, we have provided evidence that the bootstrap estimate of the standard error of $\hat{\rho}_\tau$ is consistent.

- We have completed the required computer programs to implement the asymptotic inference of $\hat{S}$ and $\hat{\rho}_\tau$, and to carry out bootstrap estimation of the standard errors of $\hat{S}$ and $\hat{\rho}_\tau$. The computer software is made available to the public via the internet.

E. REPORTABLE OUTCOMES

- **Two published articles:**

  [a] Wong, G. Y. C. and Yu, Q. Q. (1999). Generalized MLE of a joint distribution function with multivariate interval-censored data. *J. of Multi. Anal.* 69, 155-166.

  [b]. Yu, Q.Q, Wong, G.Y.C. and He, Q.M. (2000). Estimation of a joint distribution function with multivariate interval-censored data when the nonparametric MLE is not unique. *Biometrical Journal*, 42, 747-763.

- **One submitted manuscript:**

  [a]. Yu, S.H., Yu, Q.Q. and Wong, G.Y.C. (2003). Consistency of the generalized MLE of the distribution function with multivariate interval-censored data.

- **Computer programs** for asymptotic GML inferences installed at http://www.math.binghamton.edu/qyu/index/html.

- **Computer programs** for bootstrap inferences of $\hat{S}$ and $\hat{\rho}_\tau$ installed at http://www.math.binghamton.edu/qyu/index/html.

F. CONCLUSIONS

In the four years of our DOD grant, we have successfully accomplished our research objectives regarding asymptotic inferences of the GMLE $\hat{S}$ of the joint survival function for multivariate interval-censored data, and of the GMLE $\hat{\rho}_\tau$ of the correlation coefficient for a pair of correlated time-to-event variables, both of which are subject to interval censoring.

Iterative calculation to obtain $\hat{S}$ in the multivariate case can be computationally very intensive. We have implemented an efficient algorithm for this purpose. We have established consistency for $\hat{S}$ and $\hat{\rho}_\tau$ under both discrete and continuous distributional assumptions. Under discrete assumptions, we have established asymptotic normality for $\hat{S}$ and $\hat{\rho}_\tau$ so that hypothesis testing can be carried out. When the distribution function of the censoring vector is continuous, asymptotic normality is not expected for both $\hat{S}$ and $\hat{\rho}_\tau$. We have implemented a bootstrap procedure to numerically obtain asymptotic interval estimates of the parameters. We have make available to the public via the internet a set of computer programs for asymptotic GML inferences of $\hat{S}$ and $\hat{\rho}_\tau$.

The results which we have established will be useful to breast cancer researchers pursuing chemoprevention intervention trials involving multiple surrogate endpoints biomarkers, and genetic epidemiologists conducting studies on familial aggregation of breast cancer and related cancers.

G. REFERENCES

[1] Wong, G. Y. C., Bradlow, H. L., Sepkovic, D., Mehl, S., Mailman, J. and Osborne, M. P. (1997). A dose-ranging study of indole-3-carbinol for breast cancer prevention. *Journal of Cellular Biochemistry Supplements* 28/29, 111-116.

[2] Wong, G. Y. C. and Yu, Q. Q. (1999). Generalized MLE of a joint distribution function with multivariate interval-censored data. *J. of Multi. Anal.* 69, 155-166.

[3] Yu, S.H., Yu, Q.Q. and Wong, G.Y.C. (2003). Consistency of the generalized MLE of the distribution function with multivariate interval-censored data. (Submitted).

[4] Yu, Q.Q, Wong, G.Y.C. and He, Q.M. (2000). Estimation of a joint distribution function with multivariate interval-censored data when the nonparametric MLE is not unique. *Biometrical Journal*, 42, 747-763.

[5] Groeneboom, P. and Wellner, J. A. (1992). Information bounds and nonparametric maximum likelihood estimation. *Birkhäuser Verlag, Basel.*

H. APPENDICES

a. 2 published articles

30 pages

b. 1 submitted manuscripts

25 pages

Biometrical Journal **42** (2000) 6, 747–763

Estimation of a Joint Distribution Function with Multivariate Interval-Censored Data when the Nonparametric MLE is not Unique

QIQING YU

SUNY
Binghamton
USA

GEORGE Y. C. WONG

Strang Cancer Prevention Center
New York
USA

QIMEI HE

SUNY
Stony Brook
USA

Summary

A nonparametric estimator of a joint distribution function F_0 of a d-dimensional random vector with interval-censored (IC) data is the generalized maximum likelihood estimator (GMLE), where $d \geq 2$. The GMLE of F_0 with univariate IC data is uniquely defined at each follow-up time. However, this is no longer true in general with multivariate IC data as demonstrated by a data set from an eye study. How to estimate the survival function and the covariance matrix of the estimator in such a case is a new practical issue in analyzing IC data. We propose a procedure in such a situation and apply it to the data set from the eye study. Our method always results in a GMLE with a nonsingular sample information matrix. We also give a theoretical justification for such a procedure. Extension of our procedure to Cox's regression model is also mentioned.

Key words: Asymptotic normality; Consistent estimate; Multivariate survival analysis.

AMS 1991 subject classification: Primary 62G05; Secondary 62G20.

1. Introduction

Multivariate interval-censored (IC) data arise in industrial life-testing and biomedical studies. The following are examples of such data.

Example 1.1: (The Colon Cancer Study (MOERTEL et al., 1990). A national intergroup trial was conducted in the 1980's to study the drugs levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. In the study, 929 patients with stage C disease were randomly assigned to observation, levamisole alone, or levamisole combined with fluorouracil. The time to cancer recurrence and the survival time were both considered important outcome measures. The survival time was right censored. However, since most of the patients were followed up with time intervals several weeks (or months) apart, the time to cancer recurrence was only known to lie in a time interval between two follow-up times. Thus we have a bivariate random vector with one variate right censored and the other interval censored.

Example 1.2: (The Italian-American Cataract Study Group (1994)). A total of 1399 persons, between 45 and 79 years of age, who had been identified in a clinic-based case control study were enrolled in a follow-up study between 1985 and 1988. The follow-up study was designed to estimate the rate of incidence and progression of cortical, nuclear, and posterior subcapsular cataracts and to evaluate the usefulness of the Lens Opacities Classification System II in a longitudinal study. Beginning in 1989, follow-up lens photographs were taken and graded at a six-month interval. Patients might skip some visits. Data were obtained from Zeiss slit-lamp and Neitz retroillumination lens photographs at each patient's visit. Consequently, the exact time that the event of interest happened was only known to lie within the period between two consecutive visits or was right censored if by the termination of the study the event still did not happen. Hence IC data for eyes arose. Each patient had two eyes and thus bivariate IC data occurred.

Nonparametric estimation of a distribution function with univariate IC data has been studied by PETO (1973), GROENEBOOM and WELLNER (1992), and YU et al. (1998), among others. A univariate IC observation is a pair of extended real numbers L_i and R_i (*i.e.*, whose values are either real numbers or $\pm\infty$) such that $L_i \leq R_i$, $i = 1, \ldots, n$. It is one of the following 4 forms: $L_i = R_i$ (exact), $0 = L_i < R_i$ (left-censored), $L_i < R_i = \infty$ (right-censored (RC)) or $0 < L_i < R_i < \infty$ (strictly interval-censored (SIC)). A d-dimensional multivariate IC observation $(L_{i1}, R_{i1}, \ldots, L_{id}, R_{id})$ has d pairs of univariate IC observations. An observation can be viewed as a d-dimensional rectangle, say $\mathcal{I}$. In this paper we refer the univariate IC data as *univariate case 2 IC data* if the data set consists of SIC observations, and/or right-censored or left-censored observations, but not exact observations. Moreover, we refer the multivariate IC data as *multivariate case 2 IC data* if (L_{ij}, R_{ij}), $i = 1, \ldots, n$, are univariate case 2 IC data for all $j = 1, \ldots, d$. The data in Example 1.2 are bivariate case 2 IC data, but the data in Example 1.1 are not since (L_{i2}, R_{i2})'s are not univariate case 2 IC data.

Nonparametric estimation of a joint distribution function with multivariate case 2 IC data was considered by WONG and YU (1999). Under a multivariate interval censorship model and the assumption that the follow-up times take finitely many values, the problem reduces to a parametric problem of estimating a multinomial distribution. If, in addition, the GMLE is unique at follow-up times, then the generalized maximum likelihood estimator (GMLE) of a distribution function is consistent at the follow-up times and is asymptotically normally distributed. The GMLE of F_0 with univariate IC data is uniquely determined at observed follow-up times. In the multivariate case it is desirable that the GMLE of F_0 is uniquely determined at $(x_1, \ldots, x_d)$, where x_i's are observed follow-up times. However, this is not true in general. In Section 2, we present such a counter-example using data set from an eye study (LESKE et al. 1996). It presents a problem on the variance estimation with multivariate IC data since the information matrix may be singular. We shall address how to estimate F_0 and the covariance matrix of the estimator in such a case in this paper.

Multivariate right-censored (MRC) data are special cases of multivariate IC data. The GMLE with MRC data may also be not unique at follow-up times. However, it has another drawback, namely, it is not a consistent estimator of a continuous distribution function (TSAI, LEURGANS, and CROWLEY, 1986). Several consistent estimators have been proposed (see for examples, DABROWSKA (1988), PRENTICE and CAI (1992), LIN and YING (1993), and van der LAAN (1996)). These estimators are essentially unique, thus the non-uniqueness of the GMLE with MRC data has not attracted attention in the literature.

In Section 2, we propose a procedure to find, in the situation of multivariate IC data, a GMLE which always has a nonsingular sample information matrix. Thus we can use the inverse of the information matrix as an estimator of the covariance matrix of the GMLE. The theoretical justification is put in Section 3. Some detailed proofs are given in the Appendix. Section 4 is a discussion on several issues including extension of our method to Cox's regression model with multivariate IC data and covariates.

2. Method of estimation

We shall introduce the GMLE of F_0 and some notations in § 2.1, present examples of non-unique GMLEs and singular information matrices in § 2.2, and explain the procedure for estimating the variance or covariance of the GMLE in § 2.3. In § 2.4, we apply our method to a date set from an eye study (LESKE et al., 1996).

2.1 *The GMLE*

Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a d-dimensional random survival vector with a joint distribution function $F_0(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1, \ldots, x_d)$. The observable random vector is $(L_1, R_1, \ldots, L_d, R_d)$, where $L_i \leq R_i$ for all i. Suppose that $(L_{11}, R_{11}, \ldots, L_{1d}, R_{1d}), \ldots, (L_{n1}, R_{n1}, \ldots, L_{nd}, R_{nd})$ are i.i.d. copies of $(L_1, R_1, \ldots, L_d, R_d)$. Each univariate IC data (L_{ij}, R_{ij}) can be viewed as an interval I_{ij}, where $I_{ij} = \begin{cases} [L_{ij}, R_{ij}] & \text{if } L_{ij} = R_{ij}, \\ (L_{ij}, R_{ij}] & \text{if } L_{ij} < R_{ij} \end{cases}$. Thus each multivariate IC observation can be viewed as a (d-dimensional) rectangle $\mathcal{I}_i = I_{i1} \times \ldots \times I_{id}$, $i = 1, \ldots, n$. Define a *maximal intersection* (MI), A, with respect to $\mathcal{I}_i$'s to be a nonempty finite intersection of $\mathcal{I}_i$s such that $A \cap \mathcal{I}_k$ equals either $\emptyset$ or A for each k. Let $\{A_1, \ldots, A_m\}$ be the collection of all possible distinct MI's. It can be shown that the GMLE of $F_0(\boldsymbol{x})$ which maximizes the generalized likelihood function, Λ_n, must assign all probability masses $s_1, \ldots, s_m$ to the sets $A_1, \ldots, A_m$. Thus the generalized likelihood function is as follows:

$$\Lambda_n = \prod_{i=1}^{n} \mu_F(\mathcal{I}_i) = \prod_{i=1}^{n} \sum_{j=1}^{m} [\delta_{ij} s_j] , \tag{2.1}$$

where μ_F is the measure induced by an arbitrary distribution function F, $\delta_{ij} = \mathbf{1}(A_j \subset \mathcal{I}_i)$, $\mathbf{1}(\cdot)$ is the indicator function, $\mathbf{S}\,(= (s_1, \ldots, s_m)^t) \in D_s$, $\mathbf{S}^t$ is the transpose of $\mathbf{S}$, and $D_s = \{\mathbf{S}; \mathbf{S} \geq 0, \sum_{i=1}^{m} s_i = 1\}$. By $\mathbf{S} \geq 0$, we mean $s_j \geq 0$ for $j = 1, \ldots, m$. Let $\mathbf{S}_0$ be the probability mass induced by μ_{F_0}.

A GMLE of $\mathbf{S}_0$ can be obtained by the self-consistent algorithm described by Turnbull (1974) for univariate IC data as follows: Let $s_j^{(0)} = 1/m$ for $j = 1, \ldots, m$. At the h-step, $s_j^{(h)} = \sum_{i=1}^{n} \frac{1}{n} \frac{\delta_{ij} s_j^{(h-1)}}{\sum_{k=1}^{m} \delta_{ik} s_k^{(h-1)}}$, $j = 1, \ldots, m, h \geq 1$. Repeat until $s_j^{(h)}$s converge. The justification of this self-consistent algorithm for multivariate IC data is similar to that given in TURNBULL (1976). Given a GMLE $\hat{\mathbf{S}}$ of $\mathbf{S}_0$, a GMLE of $F_0(\boldsymbol{x})$ is

$$\hat{F}(\boldsymbol{x}) = \sum_{j=1}^{m} \hat{s}_j \mathbf{1}(A_j \subset [0, x_1] \times \ldots \times [0, x_d]) . \tag{2.2}$$

2.2 *Non-uniqueness of the GMLE* $\hat{\mathbf{S}}$

When $d = 1$, the GMLE $\hat{\mathbf{S}}$ of $\mathbf{S}_0$ is unique (PETO (1973)), even though the GMLE of F_0 is not unique on a non-singleton MI. Under the assumption that all the random variables are discrete and take on finitely many values, it reduces to a parametric problem of estimating a multinomial distribution (TURNBULL, 1974). Thus it is easy to show that $\hat{\mathbf{S}}$ is consistent. Furthermore, if $s_i > 0$ for all i, $(\hat{s}_1, \ldots, \hat{s}_{m-1})$ is asymptotically normally distributed. Letting

$s_m = 1 - s_1 - \ldots s_{m-1}$, an estimator of the covariance matrix of $(\hat{s}_1, \ldots, \hat{\mathbf{s}}_{m-1})$ is the inverse of the sample information matrix $-\left(\frac{\partial^2 \log \Lambda_n}{\partial s_i \partial s_j}\right)_{(m-1)\times(m-1)}\Big|_{\mathbf{s}=\hat{\mathbf{s}}}$. In application, we let $\hat{s}_1, \ldots, \hat{s}_M$ be all the nonzero elements of a GMLE $\hat{\mathbf{S}}$ obtained, denote $\mathbf{s} = (s_1, \ldots, s_{M-1})^t$, $s_M = 1 - s_1 - \ldots - s_{M-1}$ and $\hat{\mathbf{s}} = (\hat{s}_1, \ldots, \hat{s}_{M-1})^t$. Let

$$J_{\hat{\mathbf{s}}} = -\frac{\partial^2 \log \Lambda_n}{\partial \mathbf{s} \partial \mathbf{s}^t}\Big|_{\mathbf{s}=\hat{\mathbf{s}}} = \left(\sum_{h=1}^{n} \frac{(\delta_{hi} - \delta_{hM})(\delta_{hj} - \delta_{hM})}{(\mu_{\hat{F}}(\mathcal{I}_h))^2}\right)_{(M-1)\times(M-1)}. \tag{2.3}$$

$J_{\hat{\mathbf{s}}}$ is also nonsingular and $J_{\hat{\mathbf{s}}}^{-1}$ is another consistent estimator of the covariance matrix of the GMLE $\hat{\mathbf{s}}$ (see TURNBULL, 1976). In view of (2.2), $\hat{F}$ is a linear function of $\hat{\mathbf{s}}$. Thus, we can estimate the covariance of $(\hat{F}(\boldsymbol{x}), \hat{F}(\mathbf{y}))$.

When $d \geq 2$, the above arguments are no longer true. The GMLE of $\mathbf{S}_0$ may not be unique and $J_{\hat{\mathbf{s}}}$ may not be positive definite. See the following bivariate examples.

Example 2.1: Suppose that a sample of size 4 consists of observations $(L_{i1}, R_{i1}, L_{i2}, R_{i2})$, $i = 1, \ldots, 4$, which equal $(1,6,1,3)$, $(1,6,4,6)$, $(1,3,1,6)$ and $(4,6,1,6)$, respectively. Then the MI's are $A_1 = (1,3] \times (1,3]$, $A_2 = (1,3] \times (4,6]$, $A_3 = (4,6] \times (1,3]$ and $A_4 = (4,6] \times (4,6]$. $\hat{\mathbf{S}}_q = q(1/2, 0, 0, 1/2) + (1-q)$ $(0, 1/2, 1/2, 0)$, $q \in (0,1)$, are all GMLEs of $\mathbf{S}_0$.

To show that $J_{\hat{\mathbf{s}}}$ in Example 2.1 is singular, we consider the general case. Verify

$$J_{\hat{\mathbf{s}}} = U_n D U_n^t, \quad \text{where} \quad U_n = \begin{pmatrix} \delta_{11} - \delta_{1M} & \ldots & \delta_{n1} - \delta_{nM} \\ \ldots & \ldots & \ldots \\ \delta_{1(M-1)} - \delta_{1M} & \ldots & \delta_{n(M-1)} - \delta_{nM} \end{pmatrix}_{(M-1)\times n}, \tag{2.4}$$

and D is an $n \times n$ diagonal matrix with positive diagonal elements $(\mu_{\hat{F}}(\mathcal{I}_i))^{-2}$, $i = 1, \ldots, n$. Denote rank (A) the rank of a matrix A. Verify that

$$J_{\hat{\mathbf{s}}} \text{ is nonsingular if and only if rank } (U_n) = M - 1\,. \tag{2.5}$$

In view of (2.1), it is easy to show the following statement.

Proposition 1: *Let $\hat{F}$ be a GMLE of F_0. Then each solution of $\mathbf{S}$ to the equations*

$$\sum_{j=1}^{m} \delta_{ij} s_j = \mu_{\hat{F}}(\mathcal{I}_i), \; i = 1, \ldots, n, \; \sum_{j=1}^{m} s_j = 1 \quad \text{and} \quad s_j \geq 0\,, \tag{2.6}$$

is also a GMLE of $\mathbf{S}_0$.

In an obvious way, rewrite the $n+1$ equations in (2.6) as a matrix form

$$\mathcal{B}\mathbf{S} = \mathbf{p}\,, \quad \text{where} \quad \mathbf{S} \geq 0 \quad \text{and} \quad \mathbf{p} = (\mu_{\hat{F}}(\mathcal{I}_1), \ldots, \mu_{\hat{F}}(\mathcal{I}_n), 1)^t\,. \tag{2.7}$$

Remark 1: In Example 2.1, let $\hat{F}_q$ be the GMLE induced by $\hat{\mathbf{S}}_q$. Then $\mu_{\hat{F}_q}(\mathcal{I}_i) = 1/4$, $i = 1, \ldots, 4$, for all $q \in [0,1]$. This is true in general in view of Eq. (2.6). Thus the matrix D in (2.4) has the same value for all GMLEs of F_0 induced by the solutions of Eq. (2.6) and so is the vector $\mathbf{p}$ in (2.7).

Hereafter, denote $r = \text{rank}\,(\mathcal{B})$. It can be shown (see Lemma 2 in Appendix B) that

$$\text{rank}\,(U_n) \leq r - 1\,. \tag{2.8}$$

In Example 2.1 $r = 3$ and $M = 4$ for the GMLE $\hat{\mathbf{S}} = (1/4, 1/4, 1/4, 1/4)$ (see, *e.g.*, Eq. (3.3) in Section 3). Thus $\text{rank}\,(U_n) \leq 2 < 3 = M - 1$ by (2.8). Consequently, the corresponding $J_{\hat{\mathbf{S}}}$ is singular by (2.5).

In general, the number of MI's are in the order of n^d. See the following example.

Example 2.2: Assume $d = 2$. Let $\mathcal{I}_1 = (1,2] \times (0,n]$, $\mathcal{I}_2 = (3,4] \times (0,n], \ldots$, $\mathcal{I}_{n/2} = (n-1,n] \times (0,n]$, $\mathcal{I}_{(n/2)+1} = (0,n] \times (1,2]$, $\mathcal{I}_{(n/2)+2} = (0,n] \times (3,4], \ldots$, $\mathcal{I}_n = (0,n] \times (n-1,n]$, be a sample of the random rectangle $\mathcal{I}$, where n is even. Then there are $(n/2)^2$ MI's, namely, $A_i = (j, j+1] \times (k, k+1]$, where $k, j = 1, 3, 5, \ldots, n-1$. It is easy to check that there are infinitely many GMLEs of $\mathbf{S}_0$, one of them is $\hat{s}_i = (n/2)^{-2}$, $i = 1, \ldots, (n/2)^2$.

In view of the example, it is possible that $m \gg n$ for a large sample size n. In particular, it is possible that $M > n+1$. If so, $J_{\hat{\mathbf{S}}}$ is nonsingular. In fact, $\text{rank}\,(U_n) \leq r-1$ by (2.8) and thus $\text{rank}\,(U_n) < M-1$, as $r \leq \min\,\{n+1, m\} < M$ by assumptions. Consequently, $J_{\hat{\mathbf{S}}}$ is singular by (2.5). Thus, for the GMLE $\hat{\mathbf{S}}$ given above, $J_{\hat{\mathbf{S}}}$ is singular unless $n = 2$.

Our simulation experiences suggest that if $r < m$ then the self-consistent algorithm (see § 2.1) with equal initial values will result in a GMLE $\hat{\mathbf{S}}$ such that $M > r$. Hence $\text{rank}\,(U_n) \leq r-1$ $(< M-1)$ by (2.8) and thus the corresponding information matrix $J_{\hat{\mathbf{S}}}$ is singular by (2.5). Therefore, we can not estimate the covariance of $(\hat{F}(\boldsymbol{x}), \hat{F}(\mathbf{y}))$ via such $J_{\hat{\mathbf{S}}}$. If we need to make confidence statements on the estimator, such a GMLE is not desirable.

2.3 *Estimation of F_0 and the covariance matrix of the estimator*

Derive a GMLE $\hat{\mathbf{S}}$ using the self-consistent algorithm in § 2.1. If $r = m$, then $J_{\hat{\mathbf{S}}}$ is nonsingular (see Lemma 1 in Appendix B) and we use $J_{\hat{\mathbf{S}}}^{-1}$ as the estimate of the covariance matrix of $(\hat{s}_1, \ldots, \hat{s}_{M-1})$.

If $r < m$, then there are multiple solutions $\mathbf{S}$ of $\mathcal{B}\mathbf{S} = \mathbf{p}$ (Eq. (2.7)) in which $\hat{F}$ is the cdf. induced by the given $\hat{\mathbf{S}}$, and we shall look for another GMLE to replace $\hat{F}$. It can be shown (see Lemma 3 in Appendix B) that

$$\begin{aligned}&\text{there exist } r \text{ linearly independent column vectors of } \mathcal{B}\text{, say, columns } i_1, \ldots, i_r\,,\\ &\text{such that } \mathbf{c}_o = (\mathcal{V}^t\mathcal{V})^{-1}\mathcal{V}^t\,\mathbf{p} \geq 0 \;\; (\text{write } \mathbf{c}_o^t = (c_{01}, \ldots, c_{0r}))\,,\\ &\text{where } \mathcal{V} \text{ is the } (n+1) \times r \text{ matrix consisting of columns } i_1, \ldots, i_r \text{ of } \mathcal{B}\,.\end{aligned} \tag{2.9}$$

Let $\tilde{\mathbf{S}} = (\tilde{s}_1, \ldots, \tilde{s}_m)^t$, where $\tilde{s}_{i_j} = c_{0j}$, $j = 1, \ldots, r$, and $\tilde{s}_i = 0$ if $i \notin \{i_1, \ldots, i_r\}$. Verify that $\tilde{\mathbf{S}}$ satisfies (2.7) and thus by Proposition 1, $\tilde{\mathbf{S}}$ is a GMLE of $\mathbf{S}_0$. We choose such a GMLE to replace $\hat{\mathbf{S}}$. Since there are only finitely many possible combinations of r linearly independent column vectors of $\mathcal{B}$, it is easy to implement a computer algorithm to obtain $\tilde{\mathbf{S}}$ (see, for example, Remark 2 in Appendix B).

By reordering the index, without loss of generality (WLOG), we can assume

$$\tilde{\mathbf{S}}^t = (c_{01}, \ldots, c_{0r}, 0, \ldots, 0) = (\tilde{\mathbf{s}}^t, \tilde{s}_M, 0, \ldots, 0), \tag{2.10}$$

where M is the number of nonzero entries of $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{s}}$ the $(M-1) \times 1$ vector whose entries are all positive. Replacing $\hat{\mathbf{s}}$ in (2.3) by $\tilde{\mathbf{s}}$, $J_{\tilde{\mathbf{S}}}$ is nonsingular (see Lemma 1 in Appendix B). We propose to use the GMLE of F_0 corresponding to $\tilde{\mathbf{S}}$, denoted by

$$\tilde{F}(\boldsymbol{x}) = \sum_{i=1}^{M} u_i \tilde{s}_i \left(\text{and } \tilde{F}(\mathbf{y}) = \sum_{i=1}^{M} v_i \tilde{s}_i \right), \tag{2.11}$$

where $u_j = \mathbf{1}(A_j \subset [0, x_1] \times \ldots \times [0, x_d])$ (and $v_j = \mathbf{1}(A_j \subset [0, y_1] \times \ldots \times [0, y_d])$). Then the covariance of the new GMLE $\tilde{F}$ can be estimated by

$$\mathrm{Cov}\,(\tilde{F}(\boldsymbol{x}), \tilde{F}(\mathbf{y})) = (u_1 - u_M, \ldots, u_{M-1} - u_M)\, J_{\tilde{\mathbf{S}}}^{-1} (v_1 - v_M, \ldots, v_{M-1} - v_M)^t . \tag{2.12}$$

It is obvious that $\tilde{s}_j$ is not a consistent estimator of $\mu_{F_0}(A_j)$. We shall justify the above procedure in Section 3 by showing that $\tilde{\mathbf{s}}$ is asymptotic normally distributed with the asymptotic covariance estimated by $J_{\tilde{\mathbf{S}}}^{-1}$.

It is worth mentioning that in Example 2.1 $\hat{\mathbf{S}}_q$, $q \notin [0, 1]$, are also solutions of $\mathcal{B}\mathbf{S} = \mathbf{p}$, but $\hat{\mathbf{S}}_q \geq 0$ is not true. Thus they are not GMLE of $\mathbf{S}_0$. When $r < m$, if we choose an arbitrary set of r linearly independent column vectors in $\mathcal{B}$, say columns $j_1, \ldots, j_r$ such that the solution of $\mathbf{S}$ to $\mathcal{B}\mathbf{S} = \mathbf{p}$ with $s_j = 0$ if $j \notin \{j_1, \ldots, j_r\}$, then it is possible that $s_j < 0$ for some $j \in \{j_1, \ldots, j_r\}$. Thus such a solution of $\mathbf{S}$ is not desirable.

In § 2.3, we could define $M = r$ rather than define $M =$ the number of nonzero elements in $\hat{\mathbf{S}}$ obtained in (2.10). The corresponding matrix $J_{\mathbf{S}}$ is still nonsingular. However, this approach increases the dimension of $J_{\mathbf{S}}$ and thus is not desirable from a computational point of view. This is also one of the reasons that in the univariate case TURNBULL (1976) proposes to use $J_{\hat{\mathbf{S}}}$ instead of $-\left(\dfrac{\partial^2 \Lambda_n}{\partial s_i \partial s_j}\right)_{(m-1)\times(m-1)}\Bigg|_{\mathbf{S}=\hat{\mathbf{S}}}$, though both of them are nonsingular.

2.4 *An Application to an LSC study*

The following is an application of our procedure to a set of eye data from an LSC study (LESKE et al., 1996). The LSC study is an epidemiological study of the natural history of cataract similar to Example 1.2. The Leske group followed 744 participants of a case-control cataract study in a five-year period. The major aims of the study are to collect epidemiological data and to measure the growth rates (survival functions) of nuclear, cortical and posterior subcapsular opacities in a clinic-based population, to assess and compare various qualitative and quantitative methods to document changes in opacities and color, and to evaluate risk factors.

Here $\mathbf{X} = (X_1, X_2)$, where X_1 and X_2 are the time when the changes in opacities of the left and right eyes occur, respectively. The original data were recorded in the unit of days. In our analysis, we grouped the data for computational reason. Otherwise, we would end up with a large amount of MI's and thus it is difficult to compute the inverse of the information matrix even if the matrix is nonsingular. For the results in Table 1, we grouped the data in the unit of years in the following way: Let (L, R) be the original observation and (L_g, R_g) the observation after grouping. Then L_g is the largest integer that $\leq L/365$ and R_g is the smallest integer that is $\geq R/365$. We compute a GMLE of $\mathbf{S}_0$ with the grouped eye data. For this GMLE $\hat{\mathbf{S}}$, there were 27 positive entries, but the rank of the 26×26 information matrix is only 22. Thus it is singular and the GMLE of $\mathbf{S}_0$ for this data set is not unique. Using the procedure we proposed in this paper, we are able to compute the estimates of the SD of the GMLE.

In Table 1, we give the estimates of survival functions $\bar{F}(x) = P(\mathbf{X} > x)$, in the first row of each cell, and their standard deviations in the second row of each cell. Rows and columns correspond to left and right eyes, respectively. For ease in display we only give the estimate at year i (for the left eye) and year j (for the right eye).

Table 1
Estimates of $\bar{F}(i, j)$ and Their SD

year	1	2	3	4	5	6	7
1	0.968	0.911	0.880	0.855	0.815	0.797	0.761
	0.010	0.014	0.013	0.014	0.017	0.023	0.038
2	0.919	0.886	0.858	0.834	0.794	0.784	0.753
	0.015	0.015	0.014	0.015	0.018	0.024	0.039
3	0.862	0.828	0.828	0.804	0.781	0.776	0.747
	0.017	0.017	0.017	0.018	0.019	0.023	0.040
4	0.853	0.819	0.819	0.803	0.780	0.775	0.746
	0.017	0.017	0.017	0.016	0.018	0.022	0.037
5	0.819	0.786	0.786	0.770	0.770	0.765	0.737
	0.020	0.020	0.020	0.021	0.020	0.023	0.038
6	0.813	0.779	0.779	0.764	0.764	0.764	0.735
	0.024	0.024	0.024	0.023	0.023	0.023	0.038
7	0.777	0.743	0.743	0.735	0.735	0.735	0.735
	0.038	0.039	0.039	0.038	0.038	0.038	0.038

3. Theoretical issues

We shall show that under proper assumptions **all** the GMLEs are consistent on the set of all vertexes of the observed rectangles $\mathcal{I}_i$'s. Furthermore, we shall show that the GMLE we proposed is asymptotically normally distributed on the above mentioned set. For a better presentation, we put the latter proof in Appendix A.

GROENEBOOM and WELLNER (1992) formulate the univariate case 2 interval censorship model (UC2 model) for univariate case 2 data. WONG and YU (1999) formulate its natural extension, the multivariate case 2 interval censorship model (MC2 model) as follows. Suppose that the random censoring vector $(U_1, V_1, \ldots, U_d, V_d)$ and $\mathbf{X}$ are independent. The observable random vector $(L_1, R_1, \ldots, L_d, R_d)$ is generated by the following formula.

$$(L_i, R_i) = \begin{cases} (0, U_i) & \text{if} \quad X_i \leq U_i\,, \\ (U_i, V_i) & \text{if} \quad U_i < X_i \leq V_i,\ i = 1, \ldots, d\,. \\ (V_i, +\infty) & \text{if} \quad X_i > V_i\,, \end{cases} \tag{3.1}$$

The UC2 model and the MC2 model are appealing for their simplicity. However, the independence assumption between (U_i, V_i) and X_i is often not true. The reason is as follows. Univariate case 2 IC data occurred in the following situation: A patient was interviewed K times during a study period, where K may not be the same for all patients in the study. Let Y_i be the ith interview time of the patient. If the event of interest was diagnosed at time Y_i, the exact time that the event took place was only known to lie in between the two consecutive interview times Y_{i-1} and Y_i. Thus univariate IC observations can be represented by an extended random vector (L, R), where

$$(L, R) = \begin{cases} (0, Y_1) & \text{if } X \leq Y_1 (\text{left censored})\,, \\ (Y_K, +\infty) & \text{if } X > Y_K (\text{RC})\,, \\ (Y_{i-1}, Y_i) & \text{if } Y_{i-1} < X \leq Y_i \quad \text{and} \quad 2 \leq i \leq K(\text{SIC})\,. \end{cases} \tag{3.2}$$

In view of (3.1) and (3.2), we can see that (U_i, V_i) is a function of Y_j's and X_i, thus in reality, (U_i, V_i) and X_i are dependent, and it violates a key assumption in the UC2 model.

Assuming that X and $(K, \{Y_j : j \geq 1\})$ are independent, model (3.2) is called the *univariate mixed case interval censorship model* (UMC model) (SCHICK and YU, 1998). If (L_i, R_i) is from a UMC model for $i = 1, \ldots, d$, we say $(L_1, R_1, \ldots, L_d, R_d)$ is from a *multivariate mixed case interval censorship model* (MMC model). Let $\mathcal{A}_*$ be the collection of all the possible vertexes of the realizations of the random rectangle $\mathcal{I}$. The following theorem justifies all GMLEs of F_0.

Theorem 1: *Assume the MMC model and that the censoring vectors Y_js are discrete. Then each GMLE of F_0 is consistent on the set $\mathcal{A}_*$.*

The proof of the theorem is similar to the one given by YU et al. (1998b) for the UC2 model and is skipped here. Let $\mathcal{A}$ be the collection of all vertexes of the

MI's with respect to all the realizations of the random rectangle $\mathcal{I}$. Then $\mathcal{A}$ contains points $\boldsymbol{x}$, where x_i's are observed follow-up times. Note that if $\mathcal{A}_* \supset \mathcal{A}$, then it follows that the GMLE of $\mathbf{s}$ is consistent. $\mathcal{A}_* \supset \mathcal{A}$ is true for $d = 1$ but is false for $d \geq 2$. See the following example.

Example 3.1: Assume a bivariate case 2 model. Suppose that F_0 puts weights 0.4, 0.3, 0.2 and 0.1 to the points $(2,2)$, $(2,5)$, $(5,2)$ and $(5,5)$, respectively. The censoring vector $(U_1, V_1, U_2, V_2) = (1,6,1,3)$, $(1,6,4,6)$, $(1,3,1,6)$ and $(4,6,1,6)$ with probability 0.25, 0.25, 0.25 and 0.25, respectively. The possible values of (L_1, R_1, L_2, R_2) are $(1,6,1,3)$, $(1,6,4,6)$, $(1,3,1,6)$, $(4,6,1,6)$, $(1,6,0,3)$, $(1,6,4,\infty)$, $(0,3,1,6)$ and $(4,\infty,1,6)$. Denote the corresponding rectangles by $\mathcal{I}_i$, $i = 1, \ldots, 8$, respectively. Then the MI's are $A_1 = (1,3] \times (1,3]$, $A_2 = (1,3] \times (4,6]$, $A_3 = (4,6] \times (1,3]$ and $A_4 = (4,6] \times (4,6]$. The GMLE of $\mathbf{S}_0$ is not unique (Example 2.1 is a possible sample of $n = 4$) and is not consistent. However, the GMLE $\hat{F}(\boldsymbol{x})$ is uniquely defined and consistent at each $\boldsymbol{x} \in \mathcal{A}_*$, but not at (3,3), (3,4), (4,3) and (4,4), which belong to $\mathcal{A}$. In this example, rank $(\mathcal{B}) = 3$ as $\mathcal{B}\mathbf{S} = \mathbf{p}$ is equivalent to

$$
\begin{aligned}
&\mu_F(\mathcal{I}_1) = \mu_F(\mathcal{I}_5) = s_1 + s_2, \quad \mu_F(\mathcal{I}_3) = \mu_F(\mathcal{I}_7) = 1 - (s_1 + s_2)\,, \\
&\mu_F(\mathcal{I}_2) = \mu_F(\mathcal{I}_6) = s_2 + s_3, \quad \mu_F(\mathcal{I}_4) = \mu_F(\mathcal{I}_8) = 1 - (s_2 + s_3)\,, \\
&s_1 + s_2 + s_3 + s_4 = 1\,.
\end{aligned} \tag{3.3}
$$

Thus for any arbitrary sample size n, there are infinitely many GMLEs. What proposed in § 2.3 is to estimate a function F_1 such that $F_1(\boldsymbol{x}) = F_0(\boldsymbol{x})$ on $\mathcal{A}_*$ and $\mu_{F_1}(A_j) = 0$ for a fixed j, say $j = 4$. This means that we should find a GMLE with $\hat{s}_4 = 0$. Then the GMLE of $\mathbf{S}_0$ is a consistent estimator of $(\mu_{F_1}(A_1), \mu_{F_1}(A_2), \mu_{F_1}(A_3), \mu_{F_1}(A_4))$ $(= (0.5, 0.2, 0.3, 0))$, but is not a consistent estimator of $(\mu_{F_0}(A_1), \mu_{F_0}(A_2), \mu_{F_0}(A_3), \mu_{F_0}(A_4))$.

4. Discussion

4.1 *Validation of the methodology*

Is there any theoretical validation to use the inverse of the information matrix $J_{\hat{\mathbf{S}}}$ as an estimate of the covariance matrix of the GMLE $\hat{\mathbf{s}}$ (see § 2.1) with case 2 IC data? This is crucial since GROENEBOOM and WELLNER (1992) conjectured that under the assumption that all distribution functions are absolutely continuous, the GMLE of F_0 with univariate case 2 IC data is not even asymptotically normally distributed. In this regard, YU et al. (1998a, b) establish asymptotic normality results for the GMLE with univariate IC data under discrete assumptions. The current paper considers the situation that the GMLE of F_0 is not unique at discrete follow-up times. Such a discrete condition is a standard assumption in biomedical studies (see, *e.g.*, TURNBULL, 1974) and is met in most clinical studies because follow-up time is traditionally recorded in a discrete time scale such as days.

4.2 *Two types of non-uniqueness*

There are two types of non-uniqueness of $\hat{F}$ for case 2 IC data. (1) $\hat{F}$ **may** not be unique at a point in $\mathcal{A}$. (2) $\hat{F}$ is **always** not unique at the points in the interior of an MI. Example 2.1 presents an instance for both types of non-uniqueness of the GMLE. There are at least two GMLE's, say $\hat{F}$ and $\tilde{F}$, such that $\mu_{\hat{F}}(A_1) = 1/2$ but $\mu_{\tilde{F}}(A_1) = 0$ for an MI A_1, where $A_1 = (1,3] \times (1,3]$ is an MI. Thus $\hat{F}(3,3) = 1/2$ and $\tilde{F}(3,3) = 0$, where the point (3,3) is a vertex of the MI. Note that fixed $\mu_{\hat{F}}(A_i)$ for all MI's A_i, $\hat{F}$ (or $\tilde{F}$) is not uniquely determined on each A_i if $\mu_{\hat{F}}(A_i) > 0$ (or $\mu_{\tilde{F}}(A_i) > 0$). Namely, we can define $\hat{F}$ to be continuous with a density function $f = \frac{1}{8}$ on $A_1 = (1,3] \times (1,3]$, or define to be discrete with a jump $1/2$ at the point $(3,3)$ on A_1.

Only the first type of non-uniqueness causes a problem in estimating the variance of $\hat{F}(\boldsymbol{x})$. If $J_{\hat{s}}$ is singular, it indicates that we encounter the problem. In particular, the eye data in § 2.4 have the first type of non-uniqueness. Thus how to deal with such a situation is an important new issue in multivariate interval censoring as it does not occur in univariate interval censoring.

The GMLEs $\hat{F}$ obtained at the beginning of § 2.3 and $\tilde{F}$ in (2.11) are both consistent on the set $\mathcal{A}_*$. However, they can be different, even on the set $\mathcal{A}_*$. We are not aware of any proper estimator of the covariance for $\hat{F}$ if these two GMLEs are different. We propose to use $\tilde{F}$ and to use formula (2.12) as an estimate of its covariance.

4.3 *Other multivariate interval-censored data*

Multivariate right-censored data is a special case of multivariate IC data. There are other types of multivariate IC data. For instance, the data set in Example 1.1 is neither a multivariate case 2 data set nor a multivariate RC data set. Note that in Section 2, we only used the general properties of multivariate IC data and all the statements are applicable to various types of multivariate IC data. In particular, the first type of non-uniqueness also occurs in the other types of IC data. The procedure proposed in § 2 can also be applied to such data. However, the justification we make in section 3 and Appendix A will be a little bit different. To avoid complication in justification for data like multivariate RC data or the data in Example 1.1, we only consider the MMC model in Section 3.

4.4 *Multivariate right-censored (RC) data*

Even though our method is applicable to multivariate RC data, it is not a good approach since the GMLE with multivariate RC data is not a consistent estimate of a continuous F_0.

For multivariate RC data, VAN DER LAAN's (1996) modified GMLE is a more appropriate approach, since w.p.1 his estimator is consistent and is unique if the sample size is large enough. His method cannot be applied to the MMC model introduced in Section 3 since his approach takes advantage of the existence of exact observations in multivariate RC data.

HANLEY and PARNES (1983) propose an explicit estimator of the covariance of the GMLE with homogeneous multivariate RC data, that is, the right censoring vector $\mathbf{Y} = (Y_1, \dots, Y_d)$ satisfies that $Y_1 = \dots = Y_d$. Their estimate does not involve the inverse of the information matrix $J_{\hat{\mathbf{s}}}$. Multivariate case 2 IC data are unlikely homogeneous, that is, $L_{i1} = \dots = L_{id}$ and $R_{i1} = \dots = R_{id}$, $i = 1, \dots, n$. Thus this approach is not relevant in our case.

4.5 Cox's regression model

Cox's regression model is a more useful model for multivariate IC data when covariates are available. In particular, we assume the survival function $\bar{F}_0(x) = (\bar{F}_*(x))^{e^{\beta' \mathbf{z}}}$, where $\bar{F}_*$ is an unknown survival function, $\mathbf{z}$ is a covariate vector and β is a coefficient vector. It is obvious that when the covariate $\mathbf{z}$ is identical to zero, it reduces to the MMC model in Section 3. Thus the non-uniqueness of the GMLE of parameters of interest remains an obstacle in using the inverse of the information matrix as an estimate of the covariance matrix. It is conceivable that the procedure proposed in this paper can be extended to the case of Cox's regression model and always results in a positive definite information matrix.

Acknowledgement

The authors thank a referee and the editor for valuable comments. The research is partially supported by the Army grant DAMD17-99-1-9390.

Appendix

A. Justification of Formula (2.12): We shall show that $\tilde{\mathbf{s}}$ obtained in (2.10) of § 2.3 is asymptotically normally distributed under two assumptions given in due course.

Abusing notations, let $\mathcal{I}_1, \dots, \mathcal{I}_g$ be all the possible distinct realizations of the random rectangle $\mathcal{I}$, where $g < \infty$. Under this assumption, with probability one (w.p.1), for sample size n large enough, the random sample contains all $\mathcal{I}_1, \dots, \mathcal{I}_g$. WLOG, we can assume $\mathcal{I}_1, \dots, \mathcal{I}_g$ are the first g observations in the sample, and the rest are just repetitions of them. Let $A_1, \dots, A_m$ be the MI's w.r.t.

$\mathcal{I}_1, \ldots, \mathcal{I}_g$ and the ordering of A_j's corresponds to that on s_j's in (2.10). Let

$$B = \begin{pmatrix} \delta_{11} & \ldots & \delta_{1m} \\ \cdot & \ldots & \cdot \\ \delta_{g1} & \ldots & \delta_{gm} \\ 1 & \ldots & 1 \end{pmatrix}_{(g+1)\times m} \quad \text{and} \quad \mathbf{T}_{F_0} = \begin{pmatrix} \mu_F(\mathcal{I}_1) \\ \ldots \\ \mu_F(\mathcal{I}_g) \\ 1 \end{pmatrix}_{(g+1)\times 1} .$$

Denote $\gamma = \text{rank}\,(B)$. Then $\mathbf{S} = (\mu_{F_0}(A_1), \ldots, \mu_{F_0}(A_m))$ is a solution to the equation

$$B\mathbf{S} = \mathbf{T}_{F_0} \left(\text{since } \sum_{j=1}^{m} \delta_{ij} s_j = \mu_{F_0}(\mathcal{I}_i),\ i = 1, \ldots, g,\ \sum_{i=1}^{m} s_i = 1 \right), \qquad \mathbf{S} \geq 0, \tag{A.1}$$

Thus by deleting row $g+1$ through row n in Eq. (2.7), Eq. (2.7) can be simplified as

$$B\mathbf{S} = \mathbf{T}_{\hat{F}}, \text{where } \hat{F} \text{ is a GMLE and } \mathbf{S} \geq 0\,. \tag{A.2}$$

Moreover, $r = \gamma$. In view of Eq. (2.7), it follows from the theory of linear algebra that there exists a nonsingular matrix H such that $HB = \begin{pmatrix} I_\gamma & W \\ 0_1 & 0_2 \end{pmatrix}$, where I_γ is a $\gamma \times \gamma$ identity matrix, W is a $\gamma \times (m-\gamma)$ matrix, and 0_1 and 0_2 are $(g+1-\gamma) \times \gamma$ and $(g+1-\gamma) \times (m-\gamma)$ zero matrices, respectively. Then

$$\begin{pmatrix} I_\gamma & W \\ 0_1 & 0_2 \end{pmatrix} \mathbf{S} = HB\mathbf{S} = H\mathbf{T}_{\hat{F}} = \begin{pmatrix} \mathbf{c}_o \\ 0_3 \end{pmatrix}, \tag{A.3}$$

where 0_3 is a $(g+1-\gamma) \times 1$ zero matrix and $\mathbf{c}_o$ is a $\gamma \times 1$ vector. Note $\begin{pmatrix} \mathbf{c}_o \\ 0_3 \end{pmatrix} = \tilde{\mathbf{S}}$, which is the GMLE obtained in (2.10) of § 2.3, and is a solution to Eq.s (A.3) and (A.2). Hence $\mathbf{c}_o^t = (c_{01}, \ldots, c_{0\gamma})$ (see (2.10)), or $= (\tilde{\mathbf{s}}^t, \tilde{s}_M, 0, \ldots, 0)$, where $\tilde{\mathbf{s}}$ is obtained in (2.10). (A.1) yields

$$\begin{pmatrix} I_\gamma & W \\ 0_1 & 0_2 \end{pmatrix} \mathbf{S} = HB\mathbf{S} = H\mathbf{T}_{F_0} = \begin{pmatrix} \mathbf{c} \\ 0_3 \end{pmatrix}, \quad \text{where} \quad \mathbf{c} = (c_1, \ldots, c_\gamma)^t \quad \text{and} \quad \mathbf{S} \geq 0\,. \tag{A.4}$$

To justify our procedure in § 2.3, we make an additional assumption:

$$\text{If } (s_1, \ldots, s_m) \text{ is a solution to (A.1), then } \sum_{i=1}^{m} \mathbf{1}(s_j > 0) \geq \gamma\,. \tag{A.5}$$

Verify that Example 3.1 satisfies (A.5). (A.5) implies that the entries of $\mathbf{c}$ are all positive, as $\mathbf{S} = \begin{pmatrix} \mathbf{c} \\ 0_3 \end{pmatrix}$ is a solution to Eq.s (A.1) and (A.4). Since $\hat{F}$ is consistent by Theorem 1, $\mathbf{p} = T_{\hat{F}}$ converges to T_{F_0} in probability. Consequently, $\tilde{\mathbf{S}} = \begin{pmatrix} \mathbf{c}_o \\ 0_3 \end{pmatrix}$

converges to $\begin{pmatrix} \mathbf{c} \\ 0_3 \end{pmatrix}$ in probability. Thus for n large enough $M = r$ and $\mathbf{c}_o = \begin{pmatrix} \hat{\mathbf{S}} \\ \hat{s}_M \end{pmatrix}$ by (A.5).

Setting $s_{\gamma+1} = \ldots = s_m = 0$ and $s_i = \theta_i$ if $i < \gamma$, the likelihood function (2.1) becomes

$$\mathcal{L}(\theta_1, \ldots, \theta_\gamma) = \prod_{i=1}^{n} \sum_{j=1}^{\gamma} \theta_j \delta_{ij}\,, \qquad \sum_{j=1}^{\gamma} \theta_j = 1\,.$$

It is important to note that the solution $\mathbf{S}$ to Eq.s (A.2) and (A.3) with $\sum_{j=1}^{\gamma} s_j = 1$ is unique since $\begin{pmatrix} I_\gamma \\ 0_1 \end{pmatrix}$ is of rank γ. Thus the solution $\mathbf{S} = \tilde{\mathbf{S}}$ maximizes Λ_n (see (2.1)). As a consequence, $(\theta_1, \ldots, \theta_\gamma) = \mathbf{c}_o'$ maximizes $\mathcal{L}(\theta_1, \ldots, \theta_\gamma)$ under the constraint $\sum_{j=1}^{\gamma} \theta_j = 1$. Hence the MLE of $(\theta_1, \ldots, \theta_\gamma)$ is $(\tilde{\mathbf{s}}', \tilde{\mathbf{s}}_M)$ if n is large e-nough. Estimating θ_js is a parametric problem of estimating a multinomial distribution function, with parameter $(\theta_1, \ldots, \theta_\gamma)$. The MLE converges to $\mathbf{c}' > 0$. Moreover, $(\hat{\theta}_1, \ldots, \hat{\theta}_{\gamma-1})$ $(= \tilde{\mathbf{s}}')$ is asymptotically normally distributed, and a consistent estimator of its covariance matrix is $J_{\tilde{\mathbf{S}}}^{-1}$. Finally, the GMLE $\tilde{F}(\boldsymbol{x}) = \sum_{i=1}^{\gamma} u_i \hat{\theta}_i$ and $\tilde{F}(\mathbf{y}) = \sum_{i=1}^{\gamma} v_i \hat{\theta}_i$ (see (2.11)). Consequently, (2.12) gives a consistent estimator of the covariance of $\tilde{F}(\boldsymbol{x})$ and $\tilde{F}(\mathbf{y})$ under assumptions 1 and 2 (which ensures $\theta_i = c_i > 0$ for all i and $M = \gamma$ for n large enough).

B. Lemmas: We present the proofs the lemmas needed in Section 2 here.

Lemma 1: *(1) For M and $\tilde{\mathbf{s}}$ obtained in (2.10), $J_{\tilde{\mathbf{S}}}$ is nonsingular. (2) If* rank $(\mathcal{B}) = m$ *(see (2.6) and (2.7)), then $J_{\hat{\mathbf{S}}}$ (see (2.4)) is nonsingular.*

Proof: (1) Since $r = \text{rank}\,(\mathcal{B})$, there are r column vectors in $\mathcal{B}$ such that they are linearly independent. By reordering the index, WLOG, we can assume that they are the first r column vectors. Let B_1 and B_2 be $(n+1) \times M$ and $(n+1) \times (m - M)$ matrices, respectively, such that $\mathcal{B} = (B_1, B_2)$. Since the first M $(\le r)$ column vectors in $\mathcal{B}$ are linearly independent by assumption, rank $(B_1) = M$. Subtracting the last column vector of B_1 from each of the first $M - 1$ column vectors in B_1 yields

$$\begin{pmatrix} \delta_{11} - \delta_{1M} & \ldots & \delta_{1(M-1)} - \delta_{1M} & \delta_{1M} \\ \cdot & \ldots & \cdot & \cdot \\ \delta_{n1} - \delta_{1M} & \ldots & \delta_{n(M-1)} - \delta_{1M} & \delta_{nM} \\ 0 & \ldots & 0 & 1 \end{pmatrix} = \begin{pmatrix} U_n^t & \mathbf{b} \\ 0_4 & 1 \end{pmatrix}, \quad \text{where} \quad \mathbf{b} = \begin{pmatrix} \delta_{1M} \\ \cdot \\ \delta_{nM} \end{pmatrix} \tag{B.1}$$

and 0_4 is a $1 \times (M - 1)$ zero vector. Thus rank $(U_n) = M - 1$ as rank $(B_1) = M$. It follows that $J_{\tilde{\mathbf{S}}}$ is nonsingular by (2.5), which is Statement (1).

(2) Replacing M by r in the above proof results in a proof of Statement (2). □

Lemma 2: *The rank of U_n (see (2.4)) is at most $r-1$.*

Proof: In the same way as deriving (B.1), by subtracting the M-th column vector from the other column vectors in $\mathcal{B}$ (see (2.7)), the matrix $\mathcal{B}$ is equivalent to

$$\begin{pmatrix} U_n^t & \mathbf{b} & V \\ 0_4 & 1 & 0_5 \end{pmatrix}, \quad \text{where} \quad V = \begin{pmatrix} \delta_{1(M+1)} - \delta_{1M} & \dots & \delta_{1m} - \delta_{1M} \\ \dots & & \cdot \\ \delta_{n(M+1)} - \delta_{nM} & \dots & \delta_{nm} - \delta_{nM} \end{pmatrix}$$

and 0_5 is a $1 \times (n-M)$ zero vector. Thus the rank of U_n is at most $r-1$ as the M-th row vector, $(\delta_{1M}, \dots, \delta_{nM}, 1)$, is linearly independent with the remaining $m-1$ row vectors in the above $(n+1) \times m$ matrix. □

Lemma 3: *Statement (2.9) holds.*

Proof: A GMLE $\hat{\mathbf{S}}$ of $\mathbf{S}_0$ always exists and is a solution to $\mathcal{B}\mathbf{S} = \mathbf{p}$ (by Eq. (2.7)). Statement (2.9) is trivially true if rank $(\mathcal{B}) = r = m$.

Now assume rank $(\mathcal{B}) < m$. Since $\hat{\mathbf{S}}$ is a nonzero solution to $\mathcal{B}\mathbf{S} = \mathbf{p}$, there are infinitely many solutions by the theory of linear algebra. Let G be the collection of all such solutions. Note that (1) $\alpha\mathbf{S}_1 + (1-\alpha)\,\mathbf{S}_2 \in G$ for all $\mathbf{S}_1, \mathbf{S}_2 \in G$ and for all real number α; and (2) each element of $G_+ = G \cap \{\mathbf{S} \geq 0\}$ is a GMLE by Proposition 1. Thus the boundary of G_+ is not empty, *i.e.*,

$$\text{if } r < m, \text{ then } \exists\ \mathbf{S} \in G_+ \text{ such that } s_i = 0 \text{ for some } i \in \{1, \dots, m\}. \tag{B.2}$$

Deleting columns $i_1, \dots, i_j$ in the matrix $\mathcal{B}$ results in an $(n+1) \times (m-j)$ matrix $\mathcal{B}^{(i_1 \dots i_j)}$, and deleting rows $i_1, \dots, i_j$ in the column vector $\mathbf{S}$ results in an $(m-j)$ column vector $\mathbf{S}_{i_1 \dots i_j}$. By our construction, $\mathbf{S}$ is a GMLE with $s_{i_1} = \dots = s_{i_j} = 0$ iff $\mathbf{S}_{i_1 \dots i_j} \geq 0$ and $\mathbf{S}_{i_1 \dots i_j}$ is a solution to the equation $\mathcal{B}^{(i_1 \cdots i_j)}\mathbf{S}_{i_1 \dots i_j} = \mathbf{p}$. We shall show that

$$\text{if } r < m, \text{ then } \exists\, i \text{ such that rank } (\mathcal{B}^{(i)}) = r \text{ and statement (B.2) holds}. \tag{B.3}$$

Verify that if $j = 1$ then (B.3) is the same as the following statement.

$$\text{If } r < m - j + 1, \text{ then } \exists \text{ integers } i_1, \dots, i_j \text{ such that rank } (\mathcal{B}^{(i_1 \dots i_j)}) = r$$
$$\text{and } \exists \text{ a solution } \mathbf{S}_{i_1 \dots i_j} \geq 0 \text{ to the equation } \mathcal{B}^{(i_1 \cdots i_j)}\mathbf{S}_{i_1 \dots i_j} = \mathbf{p}. \tag{B.4}$$

By (B.2) and (B.3), inductively on $j \geq 1$, we can show that (B.4) holds. Now letting $j = m - r$, $\mathcal{V} = \mathcal{B}^{i_1, \dots, i_j}$ and $\mathbf{c}_o = \mathbf{S}_{i_1, \dots, i_j}$ yields Statement (2.9).

To conclude the proof, we now prove (B.3) by contradiction. Suppose that (B.3) is not true. Let $i = i_1$ satisfies (B.2). Then rank $(\mathcal{B}^{(i_1)}) = r - 1$. It follows that

$$\text{column vector } i_1 \text{ in } \mathcal{B} \text{ is linearly independent}$$
$$\text{from the rest } m - 1 \text{ column vectors}. \tag{B.5}$$

Since $B^{(i_1)}$ is an $(n+1)\times(m-1)$ matrix and rank $(B^{(i_1)}) = r-1 < m-1$, by (B.2), there is another integer i_2 $(\neq i_1)$ such that $\mathbf{S}_{i_1 i_2} \geq 0$ is a solution to the equation $B^{(i_1 i_2)}\mathbf{S}_{i_1 i_2} = \mathbf{p}$. If rank $(B^{(i_1 i_2)}) = r-1$, then $i = i_2$ must satisfy (B.3) in view of (B.5), which contradicts our assumption that (B.3) is not true. Thus rank $(B^{(i_1 i_2)}) = r-2$. Inductively on $j = 1, \ldots, r$, we would find integers $i_1, \ldots, i_j$ such that $\mathbf{S}_{i_1 j} \geq 0$ is a solution to the equation $B^{(i_1 \cdots i_j)}\mathbf{S}_{i_1 \ldots i_j} = \mathbf{p}$ and rank $(B^{(i_1 \cdots i_j)}) = r-j$. Consequently, $B^{(i_1 \cdots i_r)}$ is an $(n+1)\times(m-r)$ zero matrix as rank $(B^{(i_1 \cdots i_r)}) = r-r = 0$. It leads to $0_7 = B^{(i_1 \cdots i_r)}\mathbf{S}_{i_1 \ldots i_r} = \mathbf{p} \neq 0_7$ (due to (2.7)), where 0_7 is an $(n+1)\times 1$ zero vector. The contradiction shows that (B.3) must be true. This concludes the proof of the lemma. □

Remark 2: The proof of Lemma 3 actually provideds an explicit way to obtain Eq. (2.10). Inductively on $j = 1, \ldots, m-r$, assuming $(i_1, \ldots, i_{j-1})$ satisfies (B.4) for $j-1$, let i_j be the largest integer so that $(i_1, \ldots, i_j)$ satisfies (B.4). Let $(h_1, \ldots, h_r\} = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_{m-r}\}$. Then $\mathcal{V} = (B^{(i_1 m-r)}, \mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_{m-r}})$ is the desired matrix in (2.9).

References

DABROWSKA, D. M., 1988: Kaplan-Meier estimate on the plane. *Ann. Statist.* **16**, 1475–1489.

GROENEBOOM, P. and WELLNER, J. A., 1992: Information bounds and nonparametric maximum likelihood estimation. *Birkhäuser Verlag, Basel.*

HANLEY, J. A. and PARNES, M. N., 1983: Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics* **39**, 129–139.

The Italian-American Cataract Study Group, 1994: Incidence and progression of cortical, nuclear, and posterior subcapsular cataracts. *Amer. J. Ophthal.* **118**, *623–631.*

LESKE, M. C., CHYLACK, L. T., WU, S. Y., SCHOENFELD, E., HE, Q., FRIEND, J., WOLFE, J., and LSC Group, 1996: Incidence and progression of nuclear opacities in the longitudinal study of cataract. *Ophthalmology* **103**, 705–712.

LIN, D.Y. and YING, ZHILIANG, 1993: A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika* **80**, 573–581.

MOERTEL, C. G., FLEMING, T. R., MCDONALD, J. S., HALLER, D. G., LAURIE, J. A., GOODMAN, P. J., UNGERLEIDER, J. S., EMERSON, W. A., TORMEY, D. C., GLICK, J. H., VEEDER, M. H. and MAILLIARD, J. A., 1990: Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine* **322**, 352–358.

PETO, R., 1973: Experimental survival curves for interval-censored data. *Applied Statistics* **22**, 86–91.

PRENTICE, R. L. and CAI, J., 1992: Covariance and survival function estimation using censored multivariate failure time data. *Biometrika* **79**, 3, 495–512.

SCHICK, A. and YU, Q., 1998: Consistency of the GMLE with mixed case interval-censored data. *Scan. J. Statist.* (Accepted).

TSAI, W. Y., LEURGANS, S., and CROWLEY, J., 1986: Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.* **14**, 1351–1365.

TURNBULL, B. W., 1974: Nonparametric estimation of a survivorship function with doubly censored data. *JASA* **69**, 169–173.

TURNBULL, B. W., 1976: The empirical distribution function with arbitrary grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser.* B. **38**, 290–295.

VAN DER LAAN, M. J., 1996: Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Statist.* **24**, 596–627.

WONG, G. Y. C. and YU, Q., 1998: Generalized MLE of a joint distribution function with multivariate interval-censored data. *J. Multi. Anal.* **69**, 155–166..

YU, Q., SCHICK, A., LI, L., and WONG, G. Y. C., 1998a: Asymptotic properties of the GMLE in the case 1 interval-censorship model with discrete inspection times. *Canad. J. Statist.* **26**, 619–627.

YU, Q., SCHICK, A., LI, L., and WONG, G. Y. C., 1998b: Asymptotic properties of the GMLE of a survival function with case 2 interval-censored data. *Prob. & Statist. Let.*, 223–228.

QIQING YU
Department of Mathematics
SUNY
Binghamton, NY 13902
USA

Received, February 1999
Revised, February 2000
Accepted, April 2000

GEORGE Y. C. WONG
Strang Cancer Prevention Center
428 E 72nd Street
New York 10021
USA

Qimei He
Department of Preventive Medicine
SUNY
Stony Brook, NY 11794
USA.

Journal of Multivariate Analysis 69, 155–166 (1999)
Article ID jmva.1998.1807, available online at http://www.idealibrary.com on IDEAL®

Generalized MLE of a Joint Distribution Function with Multivariate Interval-Censored Data

George Y. C. Wong* and Qiqing Yu†

Strang Cancer Prevention Center, 428 E 72nd Street, New York 10021 and Department of Mathematical Sciences, SUNY, Binghamton

Received July 21, 1997; revised June 24, 1998

We consider the problem of estimation of a joint distribution function of a multivariate random vector with interval-censored data. The generalized maximum likelihood estimator of the distribution function is studied and its consistency and asymptotic normality are established under the case 2 multivariate interval censorship model and discrete assumptions on the censoring random vectors.

AMS 1991 subject classifications: 62G05, 62G20.

Key words and phrases: multivariate interval-censored data; asymptotic normality; asymptotic variance; consistent estimate; generalized MLE; multivariate survival analysis.

1. INTRODUCTION

We consider the estimation of a joint distribution function F_0 of a multivariate random vector $\mathbf{X} = (X_1, ..., X_d)$ which is subject to interval censoring. In interval censoring, the value of each coordinate variable X_i may not be directly observable; instead, a pair of extended real numbers L_i and R_i such that $L_i \leqslant X_i \leqslant R_i$ are always observed. The observations L_i and R_i satisfy one of the following four conditions: $L_i = R_i$ (exact), $0 = L_i < R_i$ (left censored), $L_i < R_i = \infty$ (right censored), and $0 < L_i < R_i < \infty$ (strictly interval censored). A d-dimensional interval-censored observation corresponding to $\mathbf{X}$ is represented by the $2d$-dimensional vector $(L_1, R_1, ..., L_d, R_d)$.

Multivariate interval-censored data arise in a variety of life testing situations and biomedical studies. We describe a clinical study in the

* Partially supported by Department of the Army DAMD17-94-J-4332 and DAMD 17-99-1-9390.

† Partially supported by NSF grant DMS-9402561, DAMD17-94-J-4332 and Department of the Army DAMD17-99-1-9390.

0047-259X/99 $30.00

following example that gives rise to bivariate ($d=2$) interval-censored data.

EXAMPLE 1.1 (The Italian-American Cataract Study Group (1994)). A total of 1399 persons, between 45 of 79 years of age, who had been identified in a clinic-based case control study were enrolled in a follow-up study between 1985 and 1988. The follow-up study was designed to estimate the rate of incidence and progression of cortical, nuclear, and posterior subcapsular cataracts and to evaluate the usefulness of the Lens Opacities Classification System II in a longitudinal study. Beginning in 1989, follow-up lens photographs were taken and graded at a six-month interval. Patients might skip some visits. Data were obtained from Zeiss slit-lamp and Neitz retroillumination lens photographs at each patient's visit. The exact time that the event of interest occurred was only known to lie within the period between two consecutive visits, or was right censored if by the end of the study the event still had not taken place. Consequently, bivariate interval-censored data were encountered.

At present, nonparametric estimation of a joint distribution function with multivariate interval-censored data has not been considered. A current practice is to take the midpoint of the interval (L, R) as an exact observation unless it is right censored. Then Dabrowska's (1988) Kaplan-Meier estimator on the plane or van der Laan's (1996) repaired generalized maximum likelihood estimator can be applied to such data. Another practice is to treat the right endpoints of the interval-censored data as exact observations unless they are right censored (see Samuelsen and Kongerud (1994)). However, these two practices will introduce bias in the analysis (Samuelsen and Kongerud (1994)).

Multivariate right-censored data are special cases of multivariate interval-censored data. References for nonparametric estimation of distribution functions with multivariate right-censored data can be found in Campbell (1981), Hanley and Parnes (1983), Tsai *et al.* (1986), Dabrowska (1988), Gill (1992), Prentice and Cai (1992), Lin and Ying (1993), and van der Laan (1996), etc.

Nonparametric estimation of a distribution function with univariate interval-censored data has been studied by Peto (1973), Turnbull (1976), Tsai and Crowley (1985), Chang and Yang (1987), Groeneboom and Wellner (1992), Gu and Zhang (1993), and Yu *et al* (1996 and 1998), among others.

In Section 2, we discuss generalized maximum likelihood estimation of F_0 based on multivariate interval-censored data and formulate the case 2 multivariate interval censorship model. We establish consistency of the generalized maximum likelihood estimate (GMLE) of F_0 in Section 3 and asymptotic normality of the GMLE in Section 4.

2. METHOD OF ESTIMATION

Let $\mathbf{X}=(X_1, ..., X_d)$ be a d-dimensional random survival vector with a joint distribution function $F_0(\mathbf{x})$, where $\mathbf{x}=(x_1, ..., x_d)$. The observable random vector is $(L_1, R_1, ..., L_d, R_d)$, where $L_i \leqslant R_i$ for all i. Suppose that

$$(L_{11}, R_{11}, ..., L_{1d}, R_{1d}), ..., (L_{n1}, R_{n1}, ..., L_{nd}, R_{nd})$$

are i.i.d. copies of $(L_1, R_1, ..., L_d, R_d)$. We want to estimate the joint distribution function $F_0(\mathbf{x})$ (or the survival function $S_0(\mathbf{x}) = P\{X_1 > x_1, ..., X_d > x_d\}$). Each univariate interval-censored data (L_{ij}, R_{ij}) can be viewed as an interval I_{ij}, where

$$I_{ij} = \begin{cases} [L_{ij}, R_{ij}] & \text{if } L_{ij} = R_{ij}, \\ (L_{ij}, R_{ij}] & \text{if } L_{ij} < R_{ij}; \end{cases}$$

therefore, each multivariate interval-censored observation can be viewed as a rectangular set $\mathscr{I}_i = I_{i1} \times \cdots \times I_{id}$, $i = 1, ..., n$.

Define a *maximal intersection* (MI), A, with respect to the $\mathscr{I}_i$'s to be a nonempty finite intersection of the $\mathscr{I}_i$'s such that for each i $A \cap \mathscr{I}_i = \varnothing$ or A. For example, let $\mathscr{I}_1 = (0, 2] \times (1, 3]$, $\mathscr{I}_2 = (0, 4] \times (1, 5]$, $\mathscr{I}_3 = (3, 5] \times (4, 8]$, and $\mathscr{I}_4 = (3, 5] \times (4, 8]$. Then the possible MI's are $(0, 2] \times (1, 3]$ and $(3, 4] \times (4, 5]$. Let $\{A_1, ..., A_m\}$ be the collection of all possible distinct MI's.

Using an argument similar to Hanley and Parnes (1983), it can be shown that the GMLE of $F_0(\mathbf{x})$ which maximizes the generalized likelihood function, Λ_n, must assign all the probability masses $s_1, ..., s_m$ to the sets $A_1, ..., A_m$. Thus the generalized likelihood function is as follows:

$$\Lambda_n = \prod_{i=1}^{n} \mu_F(\mathscr{I}_i) = \prod_{i=1}^{n} \left[\sum_{j=1}^{m} \mathbf{1}(A_j \subset \mathscr{I}_i)\, s_j \right], \tag{2.1}$$

where μ_F is the measure induced by a distribution function F, $\mathbf{1}(\cdot)$ is the indicator function, $\mathbf{s}$ $(=(s_1, ..., s_{m-1})^t) \in D_s$, $s_m = 1 - s_1 - \cdots - s_{m-1}$, $\mathbf{s}^t$ is the transpose of the vector $\mathbf{s}$, and $D_s = \{\mathbf{s}; s_i \geqslant 0, s_1 + \cdots + s_{m-1} \leqslant 1\}$. Denote the GMLE of $\mathbf{s}$ by $\hat{\mathbf{s}}$ and that of F_0 by $\hat{F}_n$.

The $\hat{s}_j$'s can be obtained by the self-consistent algorithm described by Turnbull (1976) for univariate interval-censored data as follows: Let $s_j^{(0)} = 1/m$ for $j = 1, ..., m$. Denote $\delta_{ij} = \mathbf{1}(A_j \subset \mathscr{I}_i)$. At the h-step, $s_j^{(h)} = \sum_{i=1}^{n} (1/n)\, (\delta_{ij} s_j^{(h-1)} / \sum_{k=1}^{m} \delta_{ik} s_k^{(h-1)})$, $j = 1, ..., m$, $h \geqslant 1$. Repeat until the s_j's converge. The justification of the convergence of this method for multivariate interval-censored data is similar to that given in Turnbull (1976) for univariate data.

Given a GMLE $\hat{\mathbf{s}}$, the GMLE of $F_0(\mathbf{x})$ is not uniquely defined on an MI unless the MI is a singleton. A GMLE of $F_0(\mathbf{x})$ can be obtained as follows:

$$\hat{F}_n(\mathbf{x}) = \sum_{A_j \subset [0, x_1] \times \cdots \times [0, x_d]} \hat{s}_j. \tag{2.2}$$

Remark 1. The GMLE of $\mathbf{s}$ may not be unique, as the following example demonstrates.

Suppose that a sample of size 4 consists of two-dimensional interval-censored observations $(1, 6, 1, 3)$, $(1, 6, 4, 6)$, $(1, 3, 1, 6)$ and $(4, 6, 1, 6)$. Then the MIs are $A_1 = (1, 3] \times (1, 3]$, $A_2 = (1, 3] \times (4, 6]$, $A_3 = (4, 6] \times (1, 3]$ and $A_4 = (4, 6] \times (4, 6]$. $(\hat{s}_1, \hat{s}_2, \hat{s}_3, \hat{s}_4) = r(1/2, 0, 0, 1/2) + (1-r)(0, 1/2, 1/2, 0)$ is a GMLE of $\mathbf{s}$, for all $r \in [0, 1]$. Thus there are infinitely many expressions for GMLE. However, $\mu_{\hat{F}_n}(\mathscr{I}_i) = 1/4$, $i = 1, ..., 4$, for all $r \in [0, 1]$.

In general, $\hat{\mathbf{s}}$ may not be consistent under discrete assumptions. However, the consistency of $\hat{F}_n$ on a certain set will not be affected (for more details, see Section 3).

The derivation of the GMLE only requires that the observations $\mathscr{I}_1, ..., \mathscr{I}_n$ are i.i.d. To derive the asymptotic properties of the GMLE, we need further assumptions on F_0 and the distribution function of $(L_1, R_1, ..., L_d, R_d)$.

A set of univariate interval-censored data are referred to as *case* 2 data if they consist of strictly interval-censored, right-censored or left-censored observations, but do not contain exact observations. For such type of data, Groeneboom and Wellner (1992) formulate the case 2 univariate interval censorship model. We consider a natural multivariate extension of the case 2 univariate interval censorship model in the following.

Suppose $(U_1, V_1, ..., U_d, V_d)$ is a random censoring vector and is independent of $\mathbf{X}$. The observable random vector $(L_1, R_1, ..., L_d, R_d)$ is generated by the following formula.

$$(L_i, R_i) = \begin{cases} (0, U_i) & \text{if } X_i \leqslant U_i, \\ (U_i, V_i) & \text{if } U_i < X_i \leqslant V_i, \quad i = 1, ..., d. \\ (V_i, +\infty) & \text{if } X_i > V_i, \end{cases}$$

We call this model a *case* 2 *multivariate interval censorship model* (C2M model). In the next two sections, we shall discuss the asymptotic properties of the GMLE under the C2M model. For ease of presentation and without loss of generality (WLOG), we assumed $d = 2$ hereafter.

3. CONSISTENCY OF GMLE

In this section, we make the following assumptions under the C2M model:

$$\text{The censoring vector } (\mathbf{U}, \mathbf{V}) \text{ is discrete.} \tag{3.1}$$

Let $\mathbf{a} = (a_1, a_2)$, $\mathbf{b} = (b_1, b_2)$, $\mathbf{U} = (U_1, U_2)$ and $\mathbf{V} = (V_1, V_2)$. Define

$$\mathscr{B} = \{(\mathbf{a}, \mathbf{b}) : g(\mathbf{a}, \mathbf{b}) > 0\}, \qquad \text{where} \quad g(\mathbf{a}, \mathbf{b}) = P(\mathbf{U} = \mathbf{a}, \mathbf{V} = \mathbf{b}),$$

Note that each point in $\mathscr{B}$ induces a grid of nine cells in R^2. Let

$$\mathscr{A}_* = \{(x_1, x_2) : x_i \in \{a_i, b_i, \pm\infty\}, i = 1, 2, (\mathbf{a}, \mathbf{b}) \in \mathscr{B}\}$$

be the set of all such grid points. We shall establish the strong consistency of the GMLE at each point in $\mathscr{A}_*$. From this we can infer the uniform strong consistency of the GMLE if F_0 is continuous and $\mathscr{A}_*$ is dense in $[0, \infty)^2$.

Let $(\mathbf{X}_i, \mathbf{U}_i, \mathbf{V}_i)$, $i = 1, ..., n$ be i.i.d. copies of $(\mathbf{X}, \mathbf{U}, \mathbf{V})$. For $(\mathbf{a}, \mathbf{b}) \in \mathscr{B}$, let

$$\begin{aligned}
&I_{11}(\mathbf{a}, \mathbf{b}) = (-\infty, a_1] \times (-\infty, a_2], \quad ..., \qquad\qquad ...,\\
&I_{21}(\mathbf{a}, \mathbf{b}) = (a_1, b_1] \times (-\infty, a_2], \quad ..., \qquad\qquad ...,\\
&I_{31}(\mathbf{a}, \mathbf{b}) = (b_1, +\infty) \times (-\infty, a_2], \quad ..., \quad I_{33}(\mathbf{a}, \mathbf{b}) = (b_1, +\infty) \times (b_2, +\infty).
\end{aligned}$$

Let $\mathscr{A}$ be the set of all vertexes of $B_1, ..., B_h$, where $B_1, ..., B_h$ are all possible MIs with respect to $I_{ij}(\mathbf{a}, \mathbf{b})$, $i, j = 1, 2, 3$, and $(\mathbf{a}, \mathbf{b}) \in \mathscr{B}$. Note that $\mathscr{A}_*$ is the set of vertexes of the rectangles $I_{ij}(\mathbf{a}, \mathbf{b})$s. Thus $\mathscr{A}_* \neq \mathscr{A}$ in general. Let

$$N_{nik}(\mathbf{a}, \mathbf{b}) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}(\mathbf{X}_j \in I_{ik}(\mathbf{a}, \mathbf{b}), \mathbf{U}_j = \mathbf{a}, \mathbf{V}_j = \mathbf{b}), \qquad i, k = 1, 2, 3.$$

Then the generalized likelihood (2.1) is equal to

$$\Lambda_n(F) = \prod_{(\mathbf{a}, \mathbf{b}) \in \mathscr{B}} \prod_{i=1}^{3} \prod_{j=1}^{3} [\mu_F(I_{ij}(\mathbf{a}, \mathbf{b}))]^{nN_{nij}(\mathbf{a}, \mathbf{b})},$$

where

$$\mu_F((c, d] \times (e, f]) = F(d, f) + F(c, e) - F(c, f) - F(d, e). \tag{3.2}$$

Moreover, the normalized generalized log-likelihood function is

$$\mathscr{L}_n(F) = \sum_{(\mathbf{a},\mathbf{b})\in\mathscr{B}} \sum_{i=1}^{3} \sum_{j=1}^{3} N_{nij}(\mathbf{a},\mathbf{b}) \ln[\mu_F(I_{ij}(\mathbf{a},\mathbf{b}))].$$

Here and below we interpret $0 \log 0 = 0$ and $\log 0 = -\infty$. For this likelihood function, we let F range over the set $\mathscr{F}^*$ of all functions F on $[-\infty, +\infty]^2$ such that

$$F(+\infty, +\infty) = 1, \tag{3.3}$$

$$F(-\infty, x) = F(x, -\infty) = 0 \qquad \text{for each } x, \tag{3.4}$$

and

$$\mu_F(I) \geqslant 0 \qquad \text{for all rectangle sets } I \text{ in } (-\infty, +\infty]^2. \tag{3.5}$$

In view of (3.2), $\Lambda_n(F)$ and $\mathscr{L}_n(F)$ depend on F only through the values of F at the points $\mathbf{x} \in \mathscr{A}_*$. Because the GMLE of F_0 is not unique, we adopt expression (2.2) for the GMLE in our proofs below.

THEOREM 1. *Under Assumption* (3.1), *the GMLE* $\hat{F}_n$ *satisfies* $\hat{F}_n(\mathbf{a}) \to F_0(\mathbf{a})$ *almost surely for all* $\mathbf{a} \in \mathscr{A}_*$.

Proof. Verify that

$$Ł(F) := E(\mathscr{L}_n(F)) = \sum_{(\mathbf{a},\mathbf{b})\in\mathscr{B}} g(\mathbf{a},\mathbf{b})\, h_{\mathbf{a},\mathbf{b}}(F) \tag{3.6}$$

with

$$h_{\mathbf{a},\mathbf{b}}(F) = \sum_{i=1}^{3} \sum_{j=1}^{3} \mu_{F_0}(I_{ij}(\mathbf{a},\mathbf{b})) \ln[\mu_F(I_{ij}(\mathbf{a},\mathbf{b}))].$$

Verify that the expression $h_{\mathbf{a},\mathbf{b}}(F)$ is maximized by a function $F \in \mathscr{F}^*$ if and only if

$$\mu_F(I_{ij}(\mathbf{a},\mathbf{b})) = \mu_{F_0}(I_{ij}(\mathbf{a},\mathbf{b})), \qquad i, j = 1, 2, 3. \tag{3.7}$$

Equations (3.2) and (3.4) imply that (3.7) is equivalent to $F(\mathbf{x}) = F_0(\mathbf{x})$ for each vertex $\mathbf{x}$ of rectangles $I_{ij}(\mathbf{a},\mathbf{b})$, $i, j = 1, 2, 3$. Thus F_0 maximizes $Ł(F)$ and any other function in $\mathscr{F}^*$ that maximizes $Ł(F)$ will coincide with F_0 on $\mathscr{A}_*$.

Note that $\mathscr{L}_n(F_0) = (1/n) \sum_{j=1}^{n} \psi(\mathbf{X}_j, \mathbf{U}_j, \mathbf{V}_j)$, where ψ is the map defined by

$$\psi(\mathbf{x}, \mathbf{a}, \mathbf{b}) = \sum_{i=1}^{3} \sum_{j=1}^{3} \mathbf{1}(\mathbf{x} \in I_{ij}(\mathbf{a},\mathbf{b})) \ln(\mu_F(I_{ij}(\mathbf{a},\mathbf{b}))).$$

Thus it follows from the SLLN and (3.2) that $\mathscr{L}_n(F_0) \to Ł(F_0)$ almost surely. By the definition of the GMLE, $\mathscr{L}_n(\hat{F}_n) \geqslant \mathscr{L}_n(F_0)$. Consequently,

$$\varliminf_{n\to\infty} \mathscr{L}_n(\hat{F}_n) \geqslant \varliminf_{n\to\infty} \mathscr{L}_n(F_0) = Ł(F_0) \text{ almost surely.}$$

Let Ω' denote the event on which $\varliminf_{n\to\infty} \mathscr{L}_n(\hat{F}_n) \geqslant Ł(F_0)$. Fix an $\omega \in \Omega'$, let $F^* \in \mathscr{F}^*$ be a limit point of $\hat{F}_{k_n}(\cdot, \omega)$ in the sense that $\hat{F}_{k_n}(\mathbf{a}, \omega) \to F^*(\mathbf{a})$ for all $\mathbf{a} \in \mathscr{A}_*$ and for some sequence $\{k_n\}$ of positive integers tending to infinity. We now show that

$$Ł(F^*) \geqslant Ł(F_0).$$

Let $t_{k_n}(\mathbf{a}, \mathbf{b})$ denote the value of the random variable $\sum_{i=1}^3 \sum_{j=1}^3 N_{k_n ij}(\mathbf{a}, \mathbf{b}) \times \ln[\mu_{\hat{F}_{k_n}}(I_{ij})]$ at the point ω. By the definition of Ω',

$$\varliminf_{n\to\infty} \sum_{(\mathbf{a},\mathbf{b})\in\mathscr{B}} t_{k_n}(\mathbf{a}, \mathbf{b}) \geqslant Ł(F_0).$$

Next, verify that

$$t_{k_n}(\mathbf{a}, \mathbf{b}) \to g(\mathbf{a}, \mathbf{b})\, h_{\mathbf{a},\mathbf{b}}(F^*)$$

for each $(\mathbf{a}, \mathbf{b}) \in \mathscr{B}$. Note also that $t_{k_n}(\mathbf{a}, \mathbf{b}) \leqslant 0$ for all $(\mathbf{a}, \mathbf{b}) \in \mathscr{B}$. From Fatou's Lemma,

$$\begin{aligned}
\varlimsup_{n\to\infty} \sum_{(\mathbf{a},\mathbf{b})\in\mathscr{B}} t_{k_n}(\mathbf{a}, \mathbf{b}) &= -\varliminf_{n\to\infty} \sum_{(\mathbf{a},\mathbf{b})\in\mathscr{B}} -t_{k_n}(\mathbf{a}, \mathbf{b}) \\
&\leqslant -\sum_{(\mathbf{a},\mathbf{b})\in\mathscr{B}} \varliminf_{n\to\infty} (-t_{k_n}(\mathbf{a}, \mathbf{b})) \\
&= \sum_{(\mathbf{a},\mathbf{b})\in\mathscr{B}} g(\mathbf{a}, \mathbf{b})\, h_{\mathbf{a},\mathbf{b}}(F^*) \\
&= Ł(F^*).
\end{aligned}$$

Combining the above yields $Ł(F_0) \leqslant Ł(F^*)$. As F_0 maximizes Ł, we conclude that $Ł(F^*) = Ł(F_0)$ and therefore $F^*(\mathbf{a}) = F_0(\mathbf{a})$ for all $\mathbf{a} \in \mathscr{A}_*$. Since ω is arbitrary and Ω' has probability one, the consistency result is thus established. ∎

If $\mathscr{A}_*$ is a finite set, then it follows from the theorem that the GMLE is uniformly strongly consistent on $\mathscr{A}_*$. For arbitrary $\mathscr{A}_*$, the uniform strong consistency of the GMLE requires additional assumptions.

THEOREM 2. *Suppose that* (3.1) *holds,* F_0 *is continuous and* $\mathscr{A}_*$ *is dense in* $[0, +\infty)^2$. *Then* $\sup_{\mathbf{x}\in\mathscr{R}^2} |\hat{F}_n(\mathbf{x}) - F_0(\mathbf{x})| \to 0$ *almost surely.*

Proof. Let $F_1, F_2, \ldots$ be functions in $\mathscr{F}^*$ such that $F_n(\mathbf{a}) \to F_0(\mathbf{a})$ for all $\mathbf{a} \in \mathscr{A}_*$. Let M be a positive integer. Since F_0 is continuous, there is a grid which partitions the space $(-\infty, +\infty]^2$ into M disjoint rectangles $I = (c, d] \times (e, f]$ with grid points (upper-right vertexes of Is) $\mathbf{x}_1, \ldots, \mathbf{x}_M$ in $(-\infty, +\infty]^2$ and $\mu_{F_0}(I) \leqslant 1/M$ for each grid cell I. The continuity of F_0 and the fact that $\mathscr{A}_*$ is dense in $[0, +\infty)^2$ imply that there are points $\mathbf{a}_1, \ldots, \mathbf{a}_M$ in $\mathscr{A}_*$ such that $|F_0(\mathbf{a}_i) - F_0(\mathbf{x}_i)| \leqslant 1/M^2$. Using this and the facts $F_0, F_n \in \mathscr{F}^*$ and that $F_0(c, e) \leqslant F_0(\mathbf{x}) \leqslant F_0(d, f)$ and $F_n(c, e) \leqslant F_n(\mathbf{x}) \leqslant F_n(d, f)$ for each $\mathbf{x} \in I$, we derive that

$$|F_n(\mathbf{x}) - F_0(\mathbf{x})| \leqslant \max_{1 \leqslant i \leqslant M} |F_n(\mathbf{a}_i) - F_0(\mathbf{a}_i)| + \frac{3}{M}, \qquad \mathbf{x} \in \mathscr{R}^2.$$

This shows that F_n converges to F_0 uniformly.

By the above, the events $\bigcap_{\mathbf{a} \in \mathscr{A}_*} \{\hat{F}_n(\mathbf{a}) \to F_0(\mathbf{a}\}$ and $\{\sup_{\mathbf{x} \in \mathscr{R}^2} |\hat{F}_n(\mathbf{x}) - F_0(\mathbf{x})| \to 0\}$ are identical and thus have probability 1 by Theorem 1. ∎

Remark 2. In the case of the bivariate right censorship model, under the assumptions in Theorem 2, it is well known that the GMLE is not a consistent estimate of a continuous F_0 (see Tsai *et al.* (1986)).

4. ASYMPTOTIC NORMALITY OF GMLE

Under the univariate case 2 interval censorship model, Groeneboom and Wellner (1992) conjecture that if the censoring distribution is continuous, then the GMLE of a continuous F_0 is not asymptotically normally distributed and the convergence rate is not in $\sqrt{n}$. Yu *et al.* (1998) prove that if the censoring vector takes on finitely many values, then under an additional assumption the GMLE is asymptotically normally distributed and the convergence rate is in $\sqrt{n}$. In the multivariate case, the situation is more complicated. In this section we shall obtain the asymptotic normality of the GMLE under the C2M model and the assumptions that

$$\mathscr{A}_* \text{ contains finitely many elements,} \tag{4.1}$$

$$\mu_{F_0}((a_1, b_1] \times (a_2, b_2]) > 0 \quad \text{if} \quad \mathbf{a}, \mathbf{b} \in \mathscr{A}_* \quad \text{and} \quad a_i < b_i, \quad i = 1, 2. \tag{4.2}$$

and

$$\mathscr{A}_* = \mathscr{A} \quad \text{(see Section 3).} \tag{4.3}$$

Note that under the current assumptions the standard method for finite parametric models can be used.

Remark 3. The GMLE of s may not be unique (see Remark 1) and Theorem 1 does not ensure the consistency of the GMLE $\hat{\mathbf{s}}$ as $\mathscr{A}$ and $\mathscr{A}_*$ are not the same in general. Note that the consistency of the GMLE $\hat{F}_n$ on $\mathscr{A}_*$ is mainly due to Eq. (3.7), since $\mathscr{A}_*$ is the set of all vertexes of the rectangles $I_{ij}(\mathbf{a}, \mathbf{b})$'s.

By Theorem 1 and (4.3), the GMLE $\hat{F}_n$ is consistent on the set $\mathscr{A}$. Since $\hat{s}_j = \mu_{\hat{F}_n}(A_j)$, where the vertexes of the MI A_j belong to $\mathscr{A}$, $\hat{\mathbf{s}}$ is consistent by (3.2).

Let $s_j^o = \mu_{F_0}(A_j)$. Then (4.2) yields $s_j^o > 0$ for all j. Verify that (3.6) yields

$$\begin{aligned}
Ł(F) &= \sum_{(\mathbf{a}, \mathbf{b}) \in \mathscr{B}} g(\mathbf{a}, \mathbf{b}) \sum_{i=1}^{3} \sum_{l=1}^{3} \sum_{k} s_k^o \mathbf{1}(A_k \subset I_{il}(\mathbf{a}, \mathbf{b})) \\
&\quad \times \ln \sum_j s_j \mathbf{1}(A_j \subset I_{il}(\mathbf{a},)) \\
&= \sum_{(\mathbf{a}, \mathbf{b}) \in \mathscr{B}} \sum_{i=1}^{3} \sum_{l=1}^{3} \left[g(\mathbf{a}, \mathbf{b}) \sum_k s_k^o \mathbf{1}(A_k \subset I_{il}(\mathbf{a}, \mathbf{b})) \right] \\
&\quad \times \ln \sum_j s_j \mathbf{1}(A_j \subset I_{il}(\mathbf{a}, \mathbf{b})). \qquad (4.4)
\end{aligned}$$

Let

$$\{I_1, ..., I_\beta\} = \{I_{ij}(\mathbf{a}, \mathbf{b}) : i, j = 1, 2, 3, (\mathbf{a}, \mathbf{b}) \in \mathscr{B}\},$$

and

$$p_h = g(\mathbf{a}, \mathbf{b}) \sum_k s_k^o \mathbf{1}(A_k \subset I_{il}(\mathbf{a}, \mathbf{b})).$$

We can rewrite (4.4) as

$$Ł(F) = \sum_{h=1}^{\beta} p_h \ln \sum_{j=1}^{m} s_j \mathbf{1}(A_j \subset I_h) = \sum_{h=1}^{\beta} p_h \ln \sum_{j=1}^{m} s_j \delta_{hj}.$$

From (4.2), $p_h > 0$, $h = 1, ..., \beta$. Set $J = -E(\partial^2 \mathscr{L}(F_0)/\partial \mathbf{s}\, \partial \mathbf{s}^t)$, where $\partial \mathscr{L}/\partial \mathbf{s}$ is an $(m-1) \times 1$ vector and $\partial^2 \mathscr{L}/\partial \mathbf{s}\, \partial \mathbf{s}^t$ is an $(m-1) \times (m-1)$ matrix. Verify that

$$J=nE\left(\frac{\partial\mathscr{L}(F_0)}{\partial\mathbf{s}}\,\frac{\partial\mathscr{L}(F_0)}{\partial\mathbf{s}^t}\right)=-\frac{\partial^2 \mathrm{Ł}}{\partial\mathbf{s}\,\partial\mathbf{s}^t}$$

$$=\left(\sum_{h=1}^{\beta} p_h \frac{(\delta_{hi}-\delta_{hm})(\delta_{hj}-\delta_{hm})}{(\sum_{k=1}^{m}\delta_{hk}s_k^o)^2}\right)_{(m-1)\times(m-1)}=UU^t,$$

where

$$U=\begin{pmatrix} \frac{(\delta_{11}-\delta_{1m})\sqrt{p_1}}{\sum_{k=1}^{m}\delta_{1k}s_k^o} & \cdots & \frac{(\delta_{\beta 1}-\delta_{\beta m})\sqrt{p_\beta}}{\sum_{k=1}^{m}\delta_{\beta k}s_k^o} \\ \cdot & \cdots & \cdot \\ \frac{(\delta_{1(m-1)}-\delta_{1m})\sqrt{p_1}}{\sum_{k=1}^{m}\delta_{1k}s_k^o} & \cdots & \frac{(\delta_{\beta(m-1)}-\delta_{\beta m})\sqrt{p_\beta}}{\sum_{k=1}^{m}\delta_{\beta k}s_k^o} \end{pmatrix}.$$

We now show that J is nonsingular. Let $\mathbf{x}_j$ be the upper-right vertex of A_j, $j=1, ..., m-1$. By reordering the I_j's, WLOG, we can assume that the upper-right vertex of I_i is equal to $\mathbf{x}_i$, $i=1, ..., m-1$. Thus $I_i \cap A_j=\varnothing$ for $j>i$, $i=1, ..., m-1$. Then the matrix U has the upper triangle matrix from

$$U=\begin{pmatrix} \frac{\sqrt{p_1}}{s_1^o} & \cdot & \cdots & \cdot & \cdots & \frac{(\delta_{\beta 1}-\delta_{\beta m})\sqrt{p_\beta}}{\sum_{k=1}^{m}\delta_{\beta k}s_k^o} \\ 0 & \frac{\sqrt{p_2}}{s_2^o+\delta_{21}s_1^o} & \cdots & \cdot & \cdots & \frac{(\delta_{\beta 2}-\delta_{\beta m})\sqrt{p_\beta}}{\sum_{k=1}^{m}\delta_{\beta k}s_k^o} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{\sqrt{p_{m-1}}}{s_{m-1}^o+\sum_{k=1}^{m-2}\delta_{(m-1)k}s_k^o} & \cdots & \frac{(\delta_{\beta(m-1)}-\delta_{\beta m})\sqrt{p_\beta}}{\sum_{k=1}^{m}\delta_{\beta k}s_k^o} \end{pmatrix}.$$

Recall $s_i^o>0$ and $p_i>0$ for $i=1, ..., m-1$. It follows that the matrix U is of full rank and $J=UU^t$ is nonsingular.

It is easy to verify that

$$\frac{\partial^2\mathscr{L}(\hat{F}_n)}{\partial\mathbf{s}\,\partial\mathbf{s}^t}\to E\left(\frac{\partial^2\mathscr{L}(F_0)}{\partial\mathbf{s}\,\partial\mathbf{s}^t}\right)=-J.$$

It thus follows that

$$\frac{\partial\mathscr{L}(\hat{F}_n)}{\partial\mathbf{s}}=\frac{\partial\mathscr{L}(F_0)}{\partial\mathbf{s}}-J\Delta_n+o_p(\|\Delta_n\|),$$

where Δ_n is the $(m-1)$-dimensional column vector with entries $\hat{s}_i - s_i^o = \mu_{\hat{F}_n}(A_i) - \mu_{F_0}(A_i)$, $i = 1, ..., m-1$. Let $\Omega_n = \{\inf_{i \leqslant m} \hat{s}_i = 0\}$. Verify that

$$0 = \frac{\partial \mathscr{L}(\hat{F}_n)}{\partial \mathbf{s}} \text{ except on the event } \Omega_n,$$

and by Theorem 1 and Assumptions (4.1) and (4.2),

$$P(\Omega_n) \to 0 \qquad \text{as} \qquad n \to \infty.$$

It follows from the CLT that $\sqrt{n}\,(\partial \mathscr{L}(F_0)/\partial \mathbf{s})$ is asymptotically normal with mean 0 and dispersion matrix J. This shows that $\Delta_n = J^{-1} \times (\partial \mathscr{L}(F_0)/\partial \mathbf{s}) + o_p(n^{-1/2})$. Thus we have the following result.

THEOREM 3. *Under Assumptions* (4.1), (4.2) *and* (4.3),

$$\sqrt{n} \begin{pmatrix} \hat{s}_1 - s_1^o \\ \vdots \\ \hat{s}_{m-1} - s_{m-1}^o \end{pmatrix}$$

is asymptotically normal with mean 0 *and dispersion matrix* J^{-1}. *A strongly consistent estimator of J is given by* $\hat{J} = -(\partial^2 \mathscr{L}(\hat{F}_n)/\partial \mathbf{s}\, \partial \mathbf{s}^t)$. *Furthermore,* $\sqrt{n}\,[\hat{F}_n(\mathbf{x}) - F_0(\mathbf{x})]$ *is asymptotically normally distributed for all* $\mathbf{x} \in \mathscr{A}_*$. *A consistent estimate of the asymptotic variance of* $\hat{F}_n(\mathbf{x})$ *is* $(1/n)\, \mathbf{c}^t \hat{J}^{-1} \mathbf{c}$, *where* $\mathbf{c}$ *is a* $(m-1) \times 1$ *vector with the ith entry* $c_i = \mathbf{1}(A_i \subset [0, x_1] \times [0, x_2])$ *unless* $F_0(\mathbf{x}) = 1$.

Under the assumptions in Theorem 3, the GMLE is also asymptotically efficient. The proof of this assertion is straightforward and is omitted.

ACKNOWLEDGMENT

We thank the referee and an editor for their invaluable suggestions and opinions.

REFERENCES

G. Campbell, Nonparametric bivariate estimation with randomly censored data, *Biometrika* **68** (1981), 417–422.

M. N. Chang and G. Yang, Strong consistency of a self-consistent estimator of the survival function with doubly censored data, *Ann. Statist.* **15** (1987), 1536–1547.

D. G. Clayton, A model of association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika* **65** (1978), 141–151.

D. G. Clayton and J. Cuzick, Multivariate generalizations of the proportional hazards model (with discussion), *J. Roy. Statist. Soc. Ser. A* **148** (1985), 82–117.

D. M. Dabrowska, Kaplan-Meier estimate on the plane, *Ann. Statist.* **16** (1988), 1475–1489.

R. Gill, Multivariate survival analysis, *Theory Prob. Appl.* **37** (1992), 18–31.

P. Groeneboom and J. A. Wellner, "Information Bounds and Nonparametric Maximum Likelihood Estimation," Birkhäuser, Basel, 1992.

M. G. Gu and C. H. Zhang, Asymptotic properties of self-consistent estimator based on doubly censored data, *Ann. Statist.* **21** (1993), 611–624.

J. A. Hanley and M. N. Parnes, Nonparametric estimation of a multivariate distribution in the presence of censoring, *Biometrics* **39** (1983), 129–139.

The Italian-American Cataract Study Group, Incidence and progression of cortical, nuclear, and posterior subcapsular cataracts, *Amer. J. Ophthal.* **118** (1994), 623–631.

E. L. Kaplan and P. Meier, Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.* **53** (1958), 457–481.

D. Y. Lin and Z. L. Ying, A simple nonparametric estimator of the bivariate survival function under univariate censoring, *Biometrika* **80** (1993), 573–581.

R. Peto, Experimental survival curves for interval-censored data, *Appl. Statist.* **22** (1973), 86–91.

R. L. Prentice and J. Cai, Covariance and survival function estimation using censored multivariate failure time data, *Biometrika* **79**(3) (1992), 495–512.

S. O. Samuelsen and J. Kongerud, Interval censoring in longitudinal data of respiratory symptoms in aluminum potroom workers: A comparison of methods, *Statist. Medicine* **13** (1994), 1771–1780.

W. Y. Tsai and J. Crowley, A large sample study of the generalized maximum likelihood estimators from incomplete data via self-consistency, *Ann. Statist.* **13** (1985), 1317–1334.

W. Y. Tsai, S. Leurgans, and J. Crowley, Nonparametric estimation of a bivariate survival function in the presence of censoring, *Ann. Statist.* **14** (1986), 1351–1365.

B. W. Turnbull, The empirical distribution function with arbitrary grouped, censored and truncated data, *J. Roy. Statist. Soc. Ser. B* **38** (1976), 290–295.

M. J. Van der Laan, Efficient estimation in the bivariate censoring model and repairing NPMLE, *Ann. Statist.* **24** (1996), 596–627.

Q. Q. Yu, A. Schick, L. L. Li, and G. Y. C. Wong, Asymptotic properties of the GMLE in the case 1 interval-censorship model with discrete inspection times, *Canad. J. Statist.*, to appear.

Q. Q. Yu, A. Schick, L. L. Li, and G. Y. C. Wong, Asymptotic properties of the GMLE of a survival function with case 2 interval-censored data, *Statist. Prob. Lett.* **37** (1998), 223–228.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium

CONSISTENCY OF THE GENERALIZED MLE OF A DISTRIBUTION FUNCTION WITH MULTIVARIATE INTERVAL-CENSORED DATA

Shaohua Yu, Qiqing Yu and George Y.C. Wong

Novartis Pharmaceuticals Corporation, New Jersey, USA

Binghamton University, New York, USA

and

Strang Cancer Prevention Center, New York, USA

Current Version: *July 21, 2003*

Short Title: *Multivariate mixed case I.C. model.*

AMS 1991 Subject Classification: *Primary 62G20; Secondary 62H12.*

Key words and phrases: *General maximum likelihood estimation, consistency, multivariate I.C. model, mixed case I.C. model, case k I.C. model.*

Abstract: We consider a mixed case multivariate interval censorship model. van der Vaart and Wellner (2000) proved strong consistency in the $L_1(\mu)$-topology of the generalized maximum likelihood estimate ($GMLE$) of the underlying distribution function, where μ is a measure derived from the joint distribution of the inspection times. However, consistency of the GMLE in other topologies has not be investigated. In this paper, we establish strong consistency of the GMLE in the topologies of weak convergence, pointwise convergence and finally uniform convergence under additional regularity conditions.

Correspondence: Qiqing Yu, Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902, USA (email: qyu@math.binghamton.edu)

1. Introduction

Interval-censored (IC) data are often encountered in longitudinal studies. The most common application is in clinical relapse follow-up studies in which the study endpoint is disease-free survival. In such a study, when a patient relapses, it is usually known that the relapse takes place between two follow-up visits, and the exact time to relapse is unknown. In statistics, we say relapse time is interval censored.

Let X denote a time-to-event variable. with distribution $F(x) = Pr(X \leq x)$, or equivalently, survival function $S(x) = 1 - F(x)$. In interval censoring, X is not observed and is known only to lie in an observable interval I with endpoints L and R. Note that (L, R) is an extended random variables, that is, $-\infty \leq L < X \leq R \leq \infty$.

The simplest model for IC data is the case 1 model (see Ayer *et al.*, 1995) in which there is only one inspection time Y, independent of X. One observes a random inspection time Y and observes whether X exceeds Y. Thus (L, R) is given by $(-\infty, Y)$ if $X \leq Y$, and (Y, ∞) otherwise.

The case 2 model (see Groeneboom & Wellner, 1992) is another model for IC data in which there are two inspection times $U < V$ that are independent of X. One observe whether an event has occurred before U, between U and V, or has not yet occurred (in other words, after V). (L, R) is defined to be the pair of endpoints of the interval among $(-\infty, U]$, $(U, V]$ or (V, ∞) that contains X. In reality, each individual in a study has K inspections and K varies within the study group. In the literature, the case 2 model has often been applied to IC data by taking U and V to be the two consecutive inspection times that $X \in (U, V]$. This treatment violates the independent assumption in the model, a key assumption in the consistency proof of the generalized mle (GMLE) of F of the case 2 model (see Groeneboom & Wellner, 1992, and Yu *et al.*, 1998).

Wellner (1995) considered a case k model in which there are k inspection times $Y_1 < ... < Y_k$, independent of X, where k is fixed. The interval among $(-\infty, Y_1]$, $(Y_1, Y_2]$, ...,$(Y_{k-1}, Y_k]$,(Y_k, ∞) that contains X is observed, and (L, R) is defined to be the endpoints of such an interval. Both case 1 model and case 2 model are special cases of the case k model. However, for $k \geq 2$, few studies satisfy the formulation of the case k model, as the number of inspection times, K, is a random variable in a study.

To accommodate the practical situation, Schick & Yu (2000) formulated a mixed case model, which assumes that the number of inspection times is random. The mixed case model can be viewed as a mixture of various case k models. The model is more realistic in practice (see, for example, the medical data in Melbye *et al.*, 1991) and has been used in Wellner and Zhang (2000), and van der Vaart and Wellner (2000).

Multivariate interval censoring involves $d \geq 2$ correlated X variables, each of which is subject to interval censoring. Under multivariate interval censoring, we consider the estimation of an underlying joint distribution function F_0 of a multivariate random vector $\mathbf{X} = (X_1, ..., X_d)'$. A multivariate interval-censored observation is d pairs of $(L_{i,\delta}, R_{i,\delta})$, where the event took place within $(L_{i,\delta}, R_{i,\delta}]$ and $0 \leq L_{i,\delta} < R_{i,\delta} \leq \infty$ for each $i = 1, ..., n$ and each $\delta = 1, ..., d$. The multivariate interval-censored data can be found in industrial life testing and medical studies. For example, in The Italian-American Cataract Study Group (1994) we can find a set of bivariate interval-censored eye data. These eye data are used to evaluated the usefulness of the Lens Opacities Classification System II. Each patient in the group is followed at a six-month interval.

Wong and Yu (1999) study a case 2 multivariate IC model and establish asymptotic properties of the GMLE of F_0. A mixed case multivariate IC model is considered in Example 1 of van der Vaart and Wellner (2000). Theorem 10 and Example 1 in van der Vaart and Wellner yield strong consistency in the $L_1(\mu)$-topology of the generalized maximum likelihood estimate of F_0, where μ is a measure derived from the joint distribution function of the inspection times. However, strong consistency in other topologies has not been addressed in the literature. In particular, uniform strong consistency results has not been established. They will be investigated in this paper.

In Section 2, we introduce the multivariate mixed case model and the consistency result in the $L_1(\mu)$-topology. In fact, a proof of the consistency result in the $L_1(\mu)$-topology is constructed by one of the authors (Yu (2000)), independently of van der Vaart and Wellner (2000). For the convenience of the reviewers of the paper, we attach the proof in the Appendix. We present strong consistency results in other topologies in Section 3. Details of some proofs are relegated to Section 4 for a better presentation.

This paper is an extension of Schick & Yu (2000). As expected, the generalization from univariate case to multivariate case is not straight forward. For instance, while the GMLE-induced measure of each maximum intersection of the observed intervals is unique in the univariate interval censoring, it is no longer so in the multivariate case (Wong & Yu, 1999). A key in the consistency proof in the univariate mixed case model is the Helly's Selection Theorem (see Rudin, 1976), which guarantees the pointwise convergence of a subsequence of distribution functions on $\mathbb{R}$. However, for higher dimensions $\mathbb{R}^d$ $(d > 1)$, Helly's Selection Theorem (Billingsley, 1968) only gives pointwise convergence on continuity points of the limiting function. Thus, topology of pointwise convergence on $\mathbb{R}^d$ is not valid. We consider the topology of pointwise convergence on a certain countable set in $\mathbb{R}^d$ and present the consistency proof in Section 4 that bypasses this difficulty.

2. Notations and preliminary results

Let $\mathbf{K} = (K_1, ..., K_d)'$ be a vector of positive random integers. K_i stands for the total number of inspection times related to X_i, $i = 1, ..., d$. Throughout the paper, we assume that $E(\prod_{i=1}^d K_i) < \infty$. This assumption is mild and generally satisfied in practice.

The *multivariate mixed case* model is formulated as follows. Conditional on $\mathbf{K} = (k_1, ..., k_d)'$, let the random vector $\mathbf{Y} = \{Y_{\delta,k_\delta,j} : \delta = 1, ..., d \text{ and } j = 1, ..., k_\delta\}$, where $k_\delta \in \mathbb{Z}^+$ and $Y_{\delta,k_\delta,1} < ... < Y_{\delta,k_\delta,k_\delta}$ are random inspection times for the δ-th coordinate. Assume that $(\mathbf{K}, \mathbf{Y})$ and $\mathbf{X}$ are independent. On the event $\{\mathbf{K} = (k_1, ..., k_d)'\}$, let $(\mathbf{L}, \mathbf{R}) = (L_1, R_1, ..., L_d, R_d)$ such that each pair (L_δ, R_δ) is from a univariate mixed case model, *i.e.*, (L_δ, R_δ) denotes the endpoints of the random interval among

$$(-\infty, Y_{\delta,k_\delta,1}],\ (Y_{\delta,k_\delta,1}, Y_{\delta,k_\delta,2}],\ ...,\ (Y_{\delta,k_\delta,k_\delta-1}, Y_{\delta,k_\delta,k_\delta}],\ (Y_{\delta,k_\delta,k_\delta}, \infty)$$

that contains X_δ, where $Y_{\delta,k_\delta,0} = -\infty$ and $Y_{\delta,k_\delta,k_\delta+1} = \infty$, $k_\delta \in \mathbb{Z}^+$.

For simplicity, assume $d = 2$. The proof for $d > 2$ is similar but much tedious. Then, $\mathbf{K} = \binom{K_1}{K_2}$, $\mathbf{X} = \binom{X_1}{X_2}$,

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \text{ where } \mathbf{Y}_\delta = \begin{pmatrix} Y_{\delta,1,1} & & \\ Y_{\delta,2,1} & Y_{\delta,2,2} & \\ Y_{\delta,3,1} & Y_{\delta,3,2} & Y_{\delta,3,3} \\ \cdots & & \end{pmatrix} \text{ for each } \delta = 1, 2.$$

Let $\mathcal{M}$ be the collection of all intervals in $\mathbb{R}$. Let $\mathcal{W}$ be the collection of all finite unions of rectangles, $A \times B$, where $A, B \in \mathcal{M}$. Obviously $\mathcal{W}$ is an algebra. We now define a set function induced by some function F, say μ_F, restricted on $\mathcal{W}$.

$$\mu_F(W) = \begin{cases} F(\binom{b}{d}) + F(\binom{a}{c}) - F(\binom{a}{d}) - F(\binom{b}{c}) & \text{if} \quad W = (a,b] \times (c,d] \\ F(\binom{a}{d}) + F(\binom{a-}{c}) - F(\binom{a-}{d}) - F(\binom{a}{c}) & \text{if} \quad W = [a,a] \times (c,d] \\ F(\binom{b}{c}) + F(\binom{a}{c-}) - F(\binom{b}{c-}) - F(\binom{a}{c}) & \text{if} \quad W = (a,b] \times [c,c] \\ F(\binom{a}{c}) + F(\binom{a}{c}-) - F(\binom{a}{c-}) - F(\binom{a-}{c}) & \text{if} \quad W = [a,a] \times [c,c], \end{cases}$$

where $F(\mathbf{x}-) = \sup\{F(\mathbf{t}) : \mathbf{t} < \mathbf{x}\}$, $F(\binom{a}{c-}) = \sup\{F(\binom{a}{t}) : t < c\}$ and $F(\binom{a-}{c}) = \sup\{F(\binom{t}{c}) : t < a\}$. Also, the notion $\mathbf{x} \leq \mathbf{y}$ $[\mathbf{x} < \mathbf{y}]$ means $x_i \leq y_i$ $[x_i < y_i]$, for all $i = 1, 2$. Let $\mathcal{F}$ be the collection of all functions from $\bar{\mathbb{R}}^2$ into $[0, 1]$ such that for each $F \in \mathcal{F}$, the following are satisfied:

1. F is nondecreasing in each variable;
2. $\mu_F(W) \geq 0$ for each $W \in \mathcal{W}$;
3. $F(\binom{\infty}{\infty}) = 1$ and $F(\binom{x}{-\infty}) = F(\binom{-\infty}{x}) = 0$, for all $x \in \mathbb{R}$.

Let $(\mathbf{L}_1, \mathbf{R}_1), ..., (\mathbf{L}_n, \mathbf{R}_n)$ be independent copies of the pair of $(\mathbf{L}, \mathbf{R})$ as defined above. Then define the generalized likelihood function

$$\Lambda_n(F) = \prod_{\eta=1}^{n} \mu_F((L_{\eta,1}, R_{\eta,1}] \times (L_{\eta,2}, R_{\eta,2}]), \quad \textit{where } F \in \mathcal{F}.$$

The normalized log-likelihood is

$$\mathcal{L}_n(F) = \frac{1}{n} \sum_{\eta=1}^{n} \log \mu_F((L_{\eta,1}, R_{\eta,1}] \times (L_{\eta,2}, R_{\eta,2}]).$$

Note that $\mathcal{L}_n(F)$ depends on F only through the values of F at the vertexes $\binom{L_{\eta,1}}{L_{\eta,2}}$, $\binom{L_{\eta,1}}{R_{\eta,2}}$, $\binom{R_{\eta,1}}{L_{\eta,2}}$ and $\binom{R_{\eta,1}}{R_{\eta,2}}$ of the half-open half-closed rectangle, for $\eta = 1, ..., n$. Thus there exist non-unique maximizers of $\mathcal{L}_n(F)$ over the set $\mathcal{F}$. However, there exists a unique maximizer $\hat{F}_n$ over $\mathcal{F}^*$, a subset of $\mathcal{F}$ containing all functions that is continuous from above and piecewise constant with possible discontinuities only at the observed values $\binom{L_{\eta,1}}{L_{\eta,2}}$, $\binom{L_{\eta,1}}{R_{\eta,2}}$, $\binom{R_{\eta,1}}{L_{\eta,2}}$ and $\binom{R_{\eta,1}}{R_{\eta,2}}$, $\eta = 1, ..., n$. We say that F is *continuous from above* at $\mathbf{x}$, if for each $\epsilon > 0$, there exists a $\delta > 0$ such that $\mathbf{x} \le \mathbf{y} < \mathbf{x} + \delta\mathbf{1}$ ($\mathbf{1}$ is the unit vector) implies that $|F(\mathbf{y}) - F(\mathbf{x})| < \epsilon$. We call this maximizer $\hat{F}_n$, the *GMLE* of F_0.

Define a measure μ on the Borel σ-field $\mathcal{B}(\mathbb{R}^2)$ such that for each $B \in \mathcal{B}(\mathbb{R}^2)$,

$$\mu(B) = \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} P\left\{\mathbf{K} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}\right\} \cdot \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} P\left\{\begin{pmatrix} Y_{1,k_1,i} \\ Y_{2,k_2,j} \end{pmatrix} \in B \,\middle|\, \mathbf{K} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}\right\}.$$

Strong consistency in $L_1(\mu)$-topology is established in the theorem below.

Theorem 2.1. $\int |\hat{F}_n - F_0| d\mu \to 0$ *a.s.*.

Recall that the assumption $E(K_1K_2) < \infty$ implies that for each $B \in \mathcal{B}(\mathbb{R}^2)$,

$$\mu(B) \le \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} k_1k_2 \cdot P\left\{\mathbf{K} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}\right\} = E(K_1K_2) < \infty.$$

A finite measure μ is vital in providing an upper bound for the integral $\int |\hat{F}_n - F_0| d\mu$, and thus a key in the consistency proof of the *GMLE* in the $L_1(\mu)$-topology. A proof of Theorem 2.1 is given in Appendix.

Remark. van der Vaart and Wellner (2000, p. 133)) point out in their Example 1 that Theorem 2.1 above is a corollary of their Theorems 9 and 10. Our proof in Appendix is different from their approach and is provided here for the convenience of readers. It can be deleted in a future revision.

The pointwise convergence for each μ-positive inspection time is obtained as a consequence of Theorem 2.1 since $\mu(\{\mathbf{a}\})|\hat{F}_n(\mathbf{a}) - F_0(\mathbf{a})| \le \int |\hat{F}_n - F_0| d\mu$ for each $\mathbf{a} \in \mathbb{R}^2$.

Corollary 2.2. $\hat{F}_n(\mathbf{a}) \to F_0(\mathbf{a})$ *a.s. for each* $\mathbf{a}$ *that satisfies* $\mu(\{\mathbf{a}\}) > 0$.

Let ν be the sum of the measures induced by the observations. For each $B \in \mathcal{B}(\mathbb{R}^2)$, $\nu(B) \le 4\mu(B)$ since $\{\binom{U_1}{U_2} \in B : U_i = L_i \text{ or } R_i, i = 1, 2\}$ is a subset of

$\bigcup_{k_1=1}^{\infty}\bigcup_{k_2=1}^{\infty}\bigcup_{i=1}^{k_1}\bigcup_{j=1}^{k_2}\left\{\binom{K_1}{K_2}=\binom{k_1}{k_2},\ \binom{Y_{1,k_1,i}}{Y_{2,k_2,j}}\in B\right\}$. Theorem 2.1 implies strong consistency for the topologies of weak convergence and the pointwise convergence for each ν-positive inspection time. Replace μ by ν, we obtain the following.

Corollary 2.3. $\int|\hat{F}_n - F_0|d\nu \to 0$ *a.s.*.

Corollary 2.4. $\hat{F}_n(\mathbf{a}) \to F_0(\mathbf{a})$ *a.s. for each* $\mathbf{a}$ *that satisfies* $\nu(\{\mathbf{a}\}) > 0$.

3. Propositions

Strong consistency in other topologies such as the topologies of weak convergence, pointwise convergence and uniform convergence are established in this section as a consequence of Theorem 2.1 with additional assumptions.

Let $\mathbf{a}, \mathbf{b}, \mathbf{x}$ be members of $\mathbb{R}^2$. For convenience, we adopt the following notations :

$$\big(\mathbf{a},\mathbf{b}\big) = \begin{cases} (a_1,b_1)\times(a_2,b_2) & \text{if } \mathbf{a}<\mathbf{b} \\ [a_1,a_1]\times(a_2,b_2) & \text{if } a_1=b_1 \text{ and } a_2<b_2 \\ (a_1,b_1)\times[a_2,a_2] & \text{if } a_2=b_2 \text{ and } a_1<b_1, \end{cases}$$

$$\big[\mathbf{a},\mathbf{b}\big] = [a_1,b_1]\times[a_2,b_2] \text{ if } \mathbf{a}\le\mathbf{b},$$

and for $\mathbf{a}<\mathbf{b}$,

$$\big[\mathbf{a},\mathbf{b}\big) = [a_1,b_1)\times[a_2,b_2)\ ,\ \big(\mathbf{a},\mathbf{b}\big] = (a_1,b_1]\times(a_2,b_2],$$

$\partial_l\big[\mathbf{a},\mathbf{b}\big]$ is the left vertical boundary $[a_1,a_1]\times[a_2,b_2]$,

$\partial_r\big[\mathbf{a},\mathbf{b}\big]$ is the right vertical boundary $[b_1,b_1]\times[a_2,b_2]$,

$\partial_u\big[\mathbf{a},\mathbf{b}\big]$ is the upper horizontal boundary $[a_1,b_1]\times[b_2,b_2]$,

$\partial_b[\mathbf{a},\mathbf{b}]$ is the bottom horizontal boundary $[a_1, b_1] \times [a_2, a_2]$, and

$\partial[\mathbf{a},\mathbf{b}] = \partial_l[\mathbf{a},\mathbf{b}] \cup \partial_r[\mathbf{a},\mathbf{b}] \cup \partial_u[\mathbf{a},\mathbf{b}] \cup \partial_b[\mathbf{a},\mathbf{b}]$;

$\mathcal{Q}$ is a square, $\boldsymbol{-}$ is a horizontal line segment and $\boldsymbol{|}$ is a vertical line segment;

$$\text{for } \Psi = \mathcal{Q},\ \boldsymbol{-} \text{ or } \boldsymbol{|},\ \mathbf{1}_\Psi = \begin{cases} (1,1)' & \text{if } \Psi = \mathcal{Q} \\ (1,0)' & \text{if } \Psi = \boldsymbol{-} \\ (0,1)' & \text{if } \Psi = \boldsymbol{|}, \end{cases}$$

$\Psi_\delta(\mathbf{x}) = (\mathbf{x}, \mathbf{x} + \delta\mathbf{1}_\Psi)$, $\Psi_{-\delta}(\mathbf{x}) = (\mathbf{x} - \delta\mathbf{1}_\Psi, \mathbf{x})$,

$\Psi_\delta[\mathbf{x}) = [\mathbf{x}, \mathbf{x} + \delta\mathbf{1}_\Psi)$, and $\Psi_{-\delta}(\mathbf{x}] = (\mathbf{x} - \delta\mathbf{1}_\Psi, \mathbf{x}]$ where $\delta > 0$;

and at last, $\mathcal{G}_\delta(\mathbf{x}), \mathrel{-\!\!+}_\delta(\mathbf{x})$ and $\dagger_\delta(\mathbf{x})$ are unions $\Psi_{-\delta}(\mathbf{x}] \cup \Psi_\delta[\mathbf{x})$,

for $\Psi = \mathcal{Q}$, $\boldsymbol{-}$ and $\boldsymbol{|}$ respectively.

We define $\mathbf{x}$ to be *a support point* of μ, if $\mu(\mathcal{G}_\delta(\mathbf{x})) > 0$ for all $\delta > 0$. Let $\mathcal{S}_\mu$ denote the set of all support points of μ. We call $\mathbf{x}$ *a horizontal support point of* μ, if $\mu(\mathrel{-\!\!+}_\delta(\mathbf{x})) > 0$ for all $\delta > 0$, and let $\mathcal{S}1_\mu$ denote the set of all horizontal support points of μ. Similarly, we define $\mathbf{x}$ to be *a vertical support point of* μ, if $\mu(\dagger_\delta(\mathbf{x})) > 0$ for all $\delta > 0$, and let $\mathcal{S}2_\mu$ to be the set of all vertical support points of μ. Define $\mathbf{x}$ to be a *regular point* of μ, if $\mu(\mathcal{Q}_{-\delta}(\mathbf{x}]) > 0$ and $\mu(\mathcal{Q}_\delta[\mathbf{x})) > 0$ for all $\delta > 0$. We say $\mathbf{x}$ is *strongly regular* with respect to μ, if $\mathbf{x}$ is a regular point of μ and $\mu(\mathcal{Q}_{-\delta}(\mathbf{x})) > 0$ for all $\delta > 0$. Notice that since $\nu \leq 4\mu$, the above concepts and the propositions and corollaries below are relevant when we replace μ by ν. We say that F *is continuous on a set* E, if for each $\mathbf{x} \in E$ and each $\epsilon > 0$, there exits a $\delta > 0$ such that $|F(\mathbf{y}) - F(\mathbf{x})| < \epsilon$, for all $\mathbf{y} \in E$ with $\rho(\mathbf{x},\mathbf{y}) < \delta$. Here $\rho(\mathbf{x},\mathbf{y}) = (\sum_{i=1}^{2}(x_i - y_i)^2)^{\frac{1}{2}}$, the Euclidean distance between x and y. Let $\mathcal{C}_{F_0}$ denote the set of all continuity points of F_0. We call $\mathbf{x}$ *a horizontal* [*vertical*] *continuity point of* F, if for each $\epsilon > 0$ there is $\delta > 0$ such that $|F(\mathbf{y}) - F(\mathbf{x})| < \epsilon$ for all $\mathbf{y} \in \mathrel{-\!\!+}_\delta(\mathbf{x})$ [$\dagger_\delta(\mathbf{x})$]. Let

$\mathcal{C}1_{F_0}$ [$\mathcal{C}2_{F_0}$] denote the set of all horizontal [vertical] continuity points of F_0. For convenience, we say F is *monotone* if a bounded function F is nondecreasing in each variable. Finally, we let $\mathcal{I}_{F_0}$ denote the set of points where F_0 is strictly increasing, *i.e.*, for each $\mathbf{x} \in \mathcal{I}_{F_0}$ and for each $\delta > 0$, $F_0(\mathbf{x} + \delta\mathbf{1}) > F_0(\mathbf{x} - \delta\mathbf{1})$. Now, consider $\Omega_\mu = \{\omega : \int_{\mathbb{R}^2} |\hat{F}_n(\mathbf{x};\omega) - F_0(\mathbf{x})| d\mu(\mathbf{x}) \to 0 \text{ as } n \to \infty\}$. By Theorem 2.1, $\mathrm{P}\{\Omega_\mu\} = 1$.

The strong consistency result for regular continuity points is given by the first proposition.

Proposition 3.1. *Suppose* $\mathbf{x} \in \mathcal{C}_{F_0}$ *is a regular point of* μ, *then* $\hat{F}_n(\mathbf{x};\omega) \to F_0(\mathbf{x})$, *for each* $\omega \in \Omega_\mu$.

The next proposition gives the weak convergence on the set of continuity points of F_0 on an open rectangle or an open line segment.

Proposition 3.2. *Let* $\mathbf{a} \leq \mathbf{b}$ *and* $\mathbf{a} \neq \mathbf{b}$. *Then the following hold :*

i. $a_1 = b_1$ *and* $(\mathbf{a}, \mathbf{b}) \subset \mathcal{S}2_\mu$ *imply that for each* $\omega \in \Omega_\mu$,
$\hat{F}_n(\mathbf{x};\omega) \to F_0(\mathbf{x})$ *for all* $\mathbf{x} \in (\mathbf{a}, \mathbf{b}) \cap \mathcal{C}2_{F_0}$;

ii. $a_2 = b_2$ *and* $(\mathbf{a}, \mathbf{b}) \subset \mathcal{S}1_\mu$ *imply that for each* $\omega \in \Omega_\mu$,
$\hat{F}_n(\mathbf{x};\omega) \to F_0(\mathbf{x})$ *for all* $\mathbf{x} \in (\mathbf{a}, \mathbf{b}) \cap \mathcal{C}1_{F_0}$;

iii. $\mathbf{a} < \mathbf{b}$ *and* $(\mathbf{a}, \mathbf{b}) \subset \mathcal{S}_\mu$ *imply that for each* $\omega \in \Omega_\mu$,
$\hat{F}_n(\mathbf{x};\omega) \to F_0(\mathbf{x})$ *for all* $\mathbf{x} \in (\mathbf{a}, \mathbf{b}) \cap \mathcal{C}_{F_0}$.

In view of Proposition 3.2, we obtain the weak convergence of the GMLE.

Proposition 3.3. *Suppose* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ *satisfy that* $F_0(\mathbf{a}) = 0$, $F_0(\mathbf{b}-) = 1$ *and* $\mu_{F_0}([\mathbf{a}, \mathbf{b}]) = 1$. *Then the following hold:*

i. $a_1 = b_1$ *and* $(\mathbf{a}, \mathbf{b}) \subset \mathcal{S}2_\mu$ *imply that for each* $\omega \in \Omega_\mu$,
$\hat{F}_n(\mathbf{x};\omega) \to F_0(\mathbf{x})$ *for all* $\mathbf{x} \in \mathcal{C}2_{F_0}$;

ii. $a_2 = b_2$ *and* $\big(\mathbf{a}, \mathbf{b}\big) \subset \mathcal{S}1_\mu$ *imply that for each* $\omega \in \Omega_\mu$,

$\hat{F}_n(\mathbf{x}; \omega) \to F_0(\mathbf{x})$ *for all* $\mathbf{x} \in \mathcal{C}1_{F_0}$;

iii. $\mathbf{a} < \mathbf{b}$, $\partial_b\big[\mathbf{a}, \mathbf{b}\big] \cup \partial_u\big[\mathbf{a}, \mathbf{b}\big] \subset \mathcal{S}1_\mu$, $\partial_l\big[\mathbf{a}, \mathbf{b}\big] \cup \partial_r\big[\mathbf{a}, \mathbf{b}\big] \subset \mathcal{S}2_\mu$ *and* $\big(\mathbf{a}, \mathbf{b}\big) \subset \mathcal{S}_\mu$ *imply that for each* $\omega \in \Omega_\mu$, $\hat{F}_n(\mathbf{x}; \omega) \to F_0(\mathbf{x})$ *for all* $\mathbf{x} \in \mathcal{C}_{F_0}$.

Proposition 3.4. *If every* $\mathbf{y} \in \mathcal{I}_{F_0}$ *is strongly regular with respect to* μ, *then for each* $\omega \in \Omega_\mu$, $\hat{F}_n(\mathbf{x}; \omega) \to F_0(\mathbf{x})$ *for all* $\mathbf{x} \in \mathcal{C}_{F_0}$.

Combining Corollary 2.1 and the above propositions, we obtain the following corollaries on pointwise convergence. Proposition 3.2 yields the pointwise convergence on an open rectangle or an open line segment. Similarly, Proposition 3.3 and Proposition 3.4 yield the pointwise convergence on the entire $\mathbb{R}^2$ plane.

Corollary 3.5. *Let* $\mathbf{a} \leq \mathbf{b}$ *and* $\mathbf{a} \neq \mathbf{b}$. *Suppose one of the assumptions listed in Proposition 3.2 is satisfied and* $\mu(\{\mathbf{y}\}) > 0$ *for each* $\mathbf{y} \in \big(\mathbf{a}, \mathbf{b}\big) \setminus \mathcal{C}_{F_0}$. *Then, for each* $\omega \in \Omega_\mu$, $\hat{F}_n(\mathbf{x}; \omega) \to F_0(\mathbf{x})$ *for all* $\mathbf{x} \in \big(\mathbf{a}, \mathbf{b}\big)$.

Corollary 3.6. *Suppose* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ *with* $F_0(\mathbf{a}) = 0$, $F_0(\mathbf{b}-) = 1$ *and* $\mu_{F_0}(\big[\mathbf{a}, \mathbf{b}\big]) = 1$ *satisfy one of the assumptions listed in Proposition 3.3. If* $\mu(\{\mathbf{y}\}) > 0$ *for each* $\mathbf{y} \in \big[\mathbf{a}, \mathbf{b}\big] \setminus \mathcal{C}_{F_0}$, *then for each* $\omega \in \Omega_\mu$, $\hat{F}_n(\mathbf{x}; \omega) \to F_0(\mathbf{x})$ *for all* $\mathbf{x} \in \mathbb{R}^2$.

Corollary 3.7. *If every* $\mathbf{y} \in \mathcal{I}_{F_0}$ *is strongly regular with respect to* μ *and* $\mu(\{\mathbf{y}\}) > 0$ *for each* $\mathbf{y} \notin \mathcal{C}_{F_0}$, *then for each* $\omega \in \Omega_\mu$, $\hat{F}_n(\mathbf{x}; \omega) \to F_0(\mathbf{x})$ *for all* $\mathbf{x} \in \mathbb{R}^2$.

We now state propositions on the uniform convergence on the entire $\mathbb{R}^2$ plane and on a closed rectangle based on the propositions and corollaries above.

Proposition 3.8. *Suppose* F_0 *is continuous. If for all* $\mathbf{a}$, $\mathbf{b} \in \mathbb{R}^2$, $\mu_{F_0}(\big(\mathbf{a}, \mathbf{b}\big)) > 0$ *implies* $\mu(\big(\mathbf{a}, \mathbf{b}\big)) > 0$, *then the GMLE is uniformly strongly consistent, i.e.,*

$$\sup_{\mathbf{x} \in \mathbb{R}^2} |\hat{F}_n(\mathbf{x}) - F_0(\mathbf{x})| \to 0 \ a.s..$$

Proposition 3.9. *Suppose that for* $\mathbf{s}, \mathbf{t} \in \mathbb{R}^2$ *satisfying* $\mathbf{s} \leq \mathbf{t}$ *and* $\mathbf{s} \neq \mathbf{t}$, *the following conditions hold :*

(a) *either* $\mu(\{\mathbf{s}\}) > 0$ *or* $F_0(\mathbf{s}) = 0$,

(b) *either* $\mu(\{\mathbf{t}\}) > 0$ *or* $F_0(\mathbf{t}-) = 1$,

(c) F_0 *is continuous on* $\big[\mathbf{s}, \mathbf{t}\big]$, *and*

(d) *for all* $\mathbf{a}, \mathbf{b} \in \big[\mathbf{s}, \mathbf{t}\big]$, $\mu_{F_0}(\big(\mathbf{a}, \mathbf{b}\big)) > 0$ *implies* $\mu(\big(\mathbf{a}, \mathbf{b}\big)) > 0$,

then the GMLE is uniformly strongly consistent on $\big[\mathbf{s}, \mathbf{t}\big]$, *i.e.,*

$$\sup_{\mathbf{x} \in \big[\mathbf{s}, \mathbf{t}\big]} |\hat{F}_n(\mathbf{x}) - F_0(\mathbf{x})| \to 0 \ a.s..$$

One may wonder whether Proposition 3.9 still holds without conditions (a) and (b). In fact, the uniform consistency results for the univariate case without these two conditions were falsely claimed in the literature (see Schick & Yu for examples). In Section 5, we will see that conditions (a) and (b) are essential for the proof.

4. Proof of Propositions

Let $\mathbb{Q}^*$ be the union of $\mathcal{A}^*$ and $\mathbb{Q}^2$, where $\mathcal{A}^* = \cup_{\rho \in \mathbb{Z}^+} \mathcal{Y}_\rho{}^2$ and $\mathbb{Q}^2$ is the set of all points in $\mathbb{R}^2$ whose coordinates are rational. Then for each $\omega \in \Omega$, there exists a subsequence $\{n'\}$ of $\{n\}$ tending to infinity such that $\hat{F}_{n'}(\mathbf{x}; \omega) \to F(\mathbf{x}; \omega)$ for all $\mathbf{x} \in \mathbb{Q}^*$, where $F \in \mathcal{F}$. To uniquely determine the F, for each $\mathbf{x} \in \mathbb{R}^2 \setminus \mathbb{Q}^*$, define $F_\omega(\mathbf{x}) = F(\mathbf{x}; \omega) = \inf\{F(\mathbf{a}; \omega) : \mathbf{a} \in \mathbb{Q}^* \text{ and } \mathbf{x} \leq \mathbf{a}\}$. Since $\hat{F}_n(\cdot; \omega)$ is a distribution function for each n and each ω, F_ω is nondecreasing in each variable and bounded by 0 and 1, obviously.

Fix an $\omega \in \Omega_\mu$. For convenience, abbreviate $\hat{F}_n(\cdot; \omega)$ by F_n, and F_ω by F. By Theorem 2.1, $\lim_{n \to \infty} \int |F_n - F_0| d\mu = \int |F - F_0| d\mu = 0$ a.s.. Let $\mathcal{D} = \{\mathbf{x} \in \mathbb{R}^2 : F(\mathbf{x}) \neq F_0(\mathbf{x})\}$. Then, $\mu(\mathcal{D}) = 0$.

PROOF OF PROPOSITION 3.1 : We shall show that if $\mathbf{x}_0 \in \mathcal{D}$ is a continuity point of F_0, then $\mathbf{x}_0$ is not regular. If $\mathcal{C}_{F_0} \cap \mathcal{D} \neq \emptyset$, there exists $\mathbf{x}_0 \in \mathcal{C}_{F_0} \cap \mathcal{D}$ such that $|F(\mathbf{x}_0) - F_0(\mathbf{x}_0)| = d > 0$. Suppose $F(\mathbf{x}_0) > F_0(\mathbf{x}_0)$. Since F_0 is continuous and monotone, there is a $\delta > 0$ such that $|F_0(\mathbf{x}) - F_0(\mathbf{x}_0)| < \frac{d}{2}$ for all $\mathbf{x} \in \mathcal{Q}_\delta[\mathbf{x}_0)$. Furthermore, $|F(\mathbf{x}) - F_0(\mathbf{x})| \geq |F(\mathbf{x}_0) - F_0(\mathbf{x})| \geq |F(\mathbf{x}_0) - F_0(\mathbf{x}_0)| - |F_0(\mathbf{x}) - F_0(\mathbf{x}_0)| > \frac{d}{2}$, for all $\mathbf{x} \in \mathcal{Q}_\delta[\mathbf{x}_0)$ by monotone property of F. Then $\mathcal{Q}_\delta[\mathbf{x}_0) \subset \mathcal{D}$ with μ-measure 0, *i.e.*, $\mathbf{x}_0$ is not regular. Similarly, if $F(\mathbf{x}_0) < F_0(\mathbf{x}_0)$, then there is a $\delta' > 0$ such that $|F_0(\mathbf{x}_0) - F_0(\mathbf{x})| < \frac{d}{2}$ for all $\mathbf{x} \in \mathcal{Q}_{-\delta'}(\mathbf{x}_0]$. Thus $\mathcal{Q}_{-\delta'}(\mathbf{x}_0]$ is in $\mathcal{D}$ with μ-measure 0, *i.e.* $\mathbf{x}_0$ is not regular. ∎

PROOF OF PROPOSITION 3.2 : We shall show that if one of the assumptions is satisfied and $\mathcal{D}$ contains a continuity point of F_0 in $\big(\mathbf{a}, \mathbf{b}\big)$, then $\mu(\mathcal{D}) > 0$, contradicting Theorem 2.1. Let $\mathcal{D}_1 = \mathcal{D} \cap \big(\mathbf{a}, \mathbf{b}\big)$. By symmetry it suffices to verify statements i and iii.

i. Assume $a_1 = b_1$ and $\big(\mathbf{a}, \mathbf{b}\big) \subset \mathcal{S}2_\mu$. Then $\mathbf{x}_0 \in \mathcal{C}2_{F_0} \cap \mathcal{D}_1$ implies that either $\mathsf{I}_{-\delta}(\mathbf{x}_0]$ or $\mathsf{I}_\delta[\mathbf{x}_0)$ is contained in $\mathcal{D}$ for some positive δ. Since $\mathbf{x}_0 \in \big(\mathbf{a}, \mathbf{b}\big) \subset \mathcal{S}2_\mu$, both $\mathsf{I}_\delta(\mathbf{x}_0)$ and $\mathsf{I}_{-\delta}(\mathbf{x}_0)$ have positive μ-measure, which leads to $\mu(\mathcal{D}_1) > 0$.

iii. Assume $\mathbf{a} < \mathbf{b}$ and $\big(\mathbf{a}, \mathbf{b}\big) \subset \mathcal{S}_\mu$. Let $\mathbf{x}_0 \in \mathcal{C}_{F_0} \cap \mathcal{D}_1$, say $|F(\mathbf{x}_0) - F_0(\mathbf{x}_0)| = d > 0$. Since F and F_0 are both monotone and $\big(\mathbf{a}, \mathbf{b}\big) \subset \mathcal{S}_\mu$. $\mathbf{x}_0$ is a continuity point of F_0, there is a $\delta > 0$ such that either $\mathcal{Q}_{-\delta}(\mathbf{x}_0]$ or $\mathcal{Q}_\delta[\mathbf{x}_0)$ is contained in $\mathcal{D}$. Since $\mathbf{x}_0$ is an interior point of $\mathcal{S}_\mu$, both $\mathcal{Q}_\delta(\mathbf{x}_0)$ and $\mathcal{Q}_{-\delta}(\mathbf{x}_0)$ have positive μ-measure. This implies $\mu(\mathcal{D}_1) > 0$. ∎

PROOF OF PROPOSITION 3.3 : Suppose that $F_0(\mathbf{a}) = 0$, $F_0(\mathbf{b}-) = 1$ and $\mu_{F_0}(\big[\mathbf{a}, \mathbf{b}\big]) = 1$, for some $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ such that $\mathbf{a} \leq \mathbf{b}$. Let $\mathcal{D}_1 = \big[\mathbf{a}, \mathbf{b}\big] \cap \mathcal{D}$. It is sufficient to show statements i and iii by symmetry.

i. Let $a_1 = b_1$ and $\big(\mathbf{a}, \mathbf{b}\big) \subset \mathcal{S}2_\mu$. Note that $\mathbf{a}, \mathbf{b} \in \mathcal{C}2_{F_0}$. Then $\mathbf{a} \notin \mathcal{D}_1$, otherwise there is a $\delta > 0$ such that $|_\delta(\mathbf{a}) \subset \mathcal{D}_1$, and thus $\mathcal{D}_1$ has positive μ-measure, a contradiction. Also, $\mathbf{b} \notin \mathcal{D}_1$, otherwise there is a $|_{-\delta}(\mathbf{b}) \subset \mathcal{D}_1$, and thus leads to the contradiction $\mu(\mathcal{D}_1) > 0$. In view of Proposition 3.2, $F(\mathbf{x}) = F_0(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{C}2_{F_0} \cap \big(\mathbf{a}, \mathbf{b}\big)$. Since $\mu_F(\big[\mathbf{a}, \mathbf{b}\big]) = \mu_{F_0}(\big[\mathbf{a}, \mathbf{b}\big]) = 1$, μ_F-measure and μ_{F_0}-measure of $\mathcal{Q}_\delta(\mathbf{x})$ $(\delta > 0)$ are 0, for each $\mathbf{x} \in \big[\mathbf{a}, \mathbf{b}\big]$. This implies that for each $\mathbf{x} \in \big[\mathbf{a}, \mathbf{b}\big]$, $F(\mathbf{y}) = F(\mathbf{x})$ and $F_0(\mathbf{y}) = F_0(\mathbf{x})$, where $\mathbf{y} \in \_\!\__\delta(\mathbf{x})$ $(\delta > 0)$. Hence, $\mathbf{x} \in \mathcal{C}2_{F_0} \cap \big[\mathbf{a}, \mathbf{b}\big]$ implies that $\mathbf{y} \in \mathcal{C}2_{F_0}$ for all $\mathbf{y} \in \_\!\__\delta(\mathbf{x})$ $(\delta > 0)$. Verify that $F(\mathbf{x}) = F_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathbb{R}^2 \setminus \big[\mathbf{a}, \infty\mathbf{1}\big)$ and $F(\mathbf{x}) = F_0(\mathbf{x}) = 1$ for all $\mathbf{x} \in \big[\mathbf{b}, \infty\mathbf{1}\big)$. Therefore, $F(\mathbf{x}) = F_0(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{C}2_{F_0}$.

iii. $\mathbf{a} < \mathbf{b}$ and $\big(\mathbf{a}, \mathbf{b}\big) \subset \mathcal{S}_\mu$. Note that $\mathbf{a}, \mathbf{b} \in \mathcal{C}_{F_0}$. Thus $\mathbf{a} \notin \mathcal{D}_1$, otherwise there is $\mathcal{Q}_\delta(\mathbf{a}) \subset \mathcal{D}_1$, and thus $\mu(\mathcal{D}_1) > 0$, a contradiction. Similarly, $\mathbf{b} \notin \mathcal{D}_1$. Notice that $\mu_{F_0}(\big[\mathbf{a}, \mathbf{b}\big]) = F_0(\mathbf{b}) + F_0(\mathbf{a}-) - F_0(\binom{a_1-}{b_2}) - F_0(\binom{b_1}{a_2-})$. For each $\mathbf{x} \in (-\infty, a_1) \times [a_2, b_2] \cup [a_1, b_1] \times (-\infty, a_2)$, $F_o(\mathbf{x}) = 0$, then by the definition of the GMLE mentioned in Section 2, $\hat{F}_n(\mathbf{x}) = 0$ and thus $F(\mathbf{x}) = F_0(\mathbf{x}) = 0$. Moreover, similar to step i., we obtain that

1. $F(\mathbf{x}) = F_0(\mathbf{x})$ for all $\mathbf{x} \in C1_{F_0} \cap (\big[\binom{a_1}{b_2}, \binom{b_1}{\infty}\big] \cup \partial_b\big[\mathbf{a}, \mathbf{b}\big])$ and for all $\mathbf{x} \in C2_{F_0} \cap (\big[\binom{b_1}{a_2}, \binom{\infty}{b_2}\big] \cup \partial_l\big[\mathbf{a}, \mathbf{b}\big])$;
2. in view of Proposition 3.2, $F(\mathbf{x}) = F_0(\mathbf{x})$ for all $\mathbf{x} \in \big(\mathbf{a}, \mathbf{b}\big) \cap \mathcal{C}_{F_0}$;
3. $F(\mathbf{x}) = F_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathbb{R}^2 \setminus \big[\mathbf{a}, \infty\mathbf{1}\big)$;
4. $F(\mathbf{x}) = F_0(\mathbf{x}) = 1$ for all $\mathbf{x} \in \big[\mathbf{b}, \infty\mathbf{1}\big)$.

Thus, $F(\mathbf{x}) = F_0(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{C}1_{F_0} \cup \mathcal{C}2_{F_0}$. ∎

PROOF OF PROPOSITION 3.4 : Let $\mathbf{x}_0 \in \mathcal{C}_{F_0}$. If $\mathbf{x}_0 \in \mathcal{I}_{F_0}$, then $\mathbf{x}_0$ is strongly regular, and hence not in $\mathcal{D}$ by Proposition 3.1. Now, suppose $\mathbf{x}_0 \notin \mathcal{I}_{F_0}$. We shall show that $\mathbf{x}_0 \notin \mathcal{D}$. Otherwise, $|F(\mathbf{x}_0) - F_0(\mathbf{x}_0)| = d > 0$. If $F(\mathbf{x}_0) > F_0(\mathbf{x}_0)$, let

$\mathbf{x} = \sup\{\mathbf{x}_0 + \delta\mathbf{1} : F_0(\mathbf{x}_0 + \delta\mathbf{1}) = F_0(\mathbf{x}_0),\ \ \delta > 0\}$. Then $\mathbf{x} \in \mathcal{I}_{F_0}$ and $\mathbf{x} = \mathbf{x}_0 + \delta_0\mathbf{1}$ for some $\delta_0 > 0$. Thus $\mu(\mathcal{Q}_{-\delta_0}(\mathbf{x})) > 0$ by assumption. Since $F(\mathbf{x}-) \geq F(\mathbf{x}_0) > F_0(\mathbf{x}_0) = F_0(\mathbf{x}-)$, $\mathcal{Q}_{-\delta_0}(\mathbf{x}) \subset \mathcal{D}$, which further implies that $\mu(\mathcal{D}) > 0$, a contradiction. On the other hand, if $F(\mathbf{x}_0) < F_0(\mathbf{x}_0)$, let $\mathbf{x} = \inf\{\mathbf{x}_0 - \delta\mathbf{1} : F_0(\mathbf{x}_0 - \delta\mathbf{1}) = F_0(\mathbf{x}_0),\ \ \delta > 0\}$. Then $\mathbf{x} \in \mathcal{I}_{F_0}$, $F(\mathbf{x}) \leq F(\mathbf{x}_0) < F_0(\mathbf{x}_0) = F_0(\mathbf{x})$, $\mathcal{Q}_{\delta_0}[\mathbf{x}) \subset \mathcal{D}$ for some $\delta_0 > 0$, and thus draw to the contradiction $\mu(\mathcal{D}) > 0$. ∎

PROOF OF PROPOSITION 3.8 : We shall show $D = \emptyset$. Otherwise, let $\mathbf{x}_0 \in \mathcal{D}$. If $F(\mathbf{x}_0) - F_0(\mathbf{x}_0) = d > 0$, let $\mathbf{x} = \sup\{\mathbf{x}_0 + \delta\mathbf{1} : F_0\big(\mathbf{x}_0 + \delta\mathbf{1}\big) = F_0(\mathbf{x}_0),\ \ \delta > 0\}$. Then $\mathbf{x} \in \mathcal{I}_{F_0}$. Since F_0 is continuous, there is a positive δ_0 such that $F_0(\mathbf{x} + \delta_0\mathbf{1}) - F_0(\mathbf{x}) < \frac{d}{2}$. Then $\mu_{F_0}(\big(\mathbf{x}, \mathbf{x} + \delta_0\mathbf{1}\big)) > 0$ and $\big(\mathbf{x}, \mathbf{x} + \delta_0\mathbf{1}\big) \subset \mathcal{D}$, which imply that $\mu(\mathcal{D}) > 0$, contradicting Theorem 2.1. The same contradiction can be reached similarly for the case $F(\mathbf{x}_0) - F_0(\mathbf{x}_0) = d < 0$. Thus $\mathcal{D} = \emptyset$ and F_n pointwisely converges to F_0.

Let $\epsilon > 0$ and $\mathbf{x}_0 \in \mathbb{R}^2$. By continuity and monotonicity of F_0, we can choose finitely many quantiles $\{a_0, a_1, ..., a_\alpha\}$ and $\{b_0, b_1, ..., b_\beta\}$ such that $a_0 = b_0 = -\infty$, $F_0(\binom{a_i}{\infty}) - F_0(\binom{a_{i-1}}{\infty}) < \epsilon$ for each $i = 1, ..., \alpha$, and $F_0(\binom{\infty}{b_j}) - F_0(\binom{\infty}{b_{j-1}}) < \epsilon$ for each $j = 1, ..., \beta$. Then there exists N such that $|F_n(\binom{a_i}{b_j}) - F_0(\binom{a_i}{b_j})| < \epsilon$ for all $i = 0, ..., \alpha$, $j = 0, ..., \beta$, and all $n > N$. Note that $\mathbf{x}_0 \in \big(\binom{a_i}{b_j}, \binom{a_{i+1}}{b_{j+1}}\big]$ for some i, j. Then $|F_0(\mathbf{y}) - F_0(\mathbf{x}_0)| < \epsilon$ for all $\mathbf{y} \in \big(\binom{a_i}{b_j}, \binom{a_{i+1}}{b_{j+1}}\big]$. Moreover,

$$
\begin{aligned}
|F_n(\begin{pmatrix} a_{i+1} \\ b_{j+1} \end{pmatrix}) - F_m(\begin{pmatrix} a_i \\ b_j \end{pmatrix})| &\leq |F_n(\begin{pmatrix} a_{i+1} \\ b_{j+1} \end{pmatrix}) - F_0(\begin{pmatrix} a_{i+1} \\ b_{j+1} \end{pmatrix})| + |F_0(\begin{pmatrix} a_{i+1} \\ b_{j+1} \end{pmatrix}) - F_0(\mathbf{x}_0)| \\
&\quad + |F_0(\mathbf{x}_0) - F_0(\begin{pmatrix} a_i \\ b_j \end{pmatrix})| + |F_0(\begin{pmatrix} a_i \\ b_j \end{pmatrix}) - F_m(\begin{pmatrix} a_i \\ b_j \end{pmatrix})| \\
&\leq 4\epsilon, \ \text{where } n, m > N.
\end{aligned}
$$

Combining with the monotonicity of F_n, we obtain

$$
\begin{aligned}
&|F_n(\mathbf{x}_0) - F_m(\mathbf{x}_0)| \\
&\leq \left|F_n(\mathbf{x}_0) - F_n\left(\begin{pmatrix} a_i \\ b_j \end{pmatrix}\right)\right| + \left|F_n\left(\begin{pmatrix} a_i \\ b_j \end{pmatrix}\right) - F_m\left(\begin{pmatrix} a_{i+1} \\ b_{j+1} \end{pmatrix}\right)\right| + \left|F_m\left(\begin{pmatrix} a_{i+1} \\ b_{j+1} \end{pmatrix}\right) - F_m(\mathbf{x}_0)\right| \\
&\leq \left|F_n\left(\begin{pmatrix} a_{i+1} \\ b_{j+1} \end{pmatrix}\right) - F_n\left(\begin{pmatrix} a_i \\ b_j \end{pmatrix}\right)\right| + \left|F_n\left(\begin{pmatrix} a_i \\ b_j \end{pmatrix}\right) - F_m\left(\begin{pmatrix} a_{i+1} \\ b_{j+1} \end{pmatrix}\right)\right| \\
&\qquad\qquad + \left|F_m\left(\begin{pmatrix} a_{i+1} \\ b_{j+1} \end{pmatrix}\right) - F_m\left(\begin{pmatrix} a_i \\ b_j \end{pmatrix}\right)\right| \\
&\leq 12\epsilon, \text{ for all } n, m > N.
\end{aligned}
\tag{4.1}
$$

Since $\mathbf{x}_0$ arbitrary, F_n converges uniformly to F_0 on $\mathbb{R}^2$. ∎

PROOF OF PROPOSITION 3.9: *WLOG*, assume $\mathbf{s} < \mathbf{t}$. First consider the case when $\mu(\{\mathbf{s}\}) > 0$ and $F_0(\mathbf{t}-) = 1$. By Corollary 2.2, $F(\mathbf{s}) = F_0(\mathbf{s})$. If $F_0(\mathbf{s}) = 1$, we are done. Now, assume $F_0(\mathbf{s}) < 1$. We shall show $\mathcal{D}_1 = \mathcal{D} \cap [\mathbf{s}, \mathbf{t}] = \emptyset$ in three steps.

(1) $\mathbf{t} \notin \mathcal{D}_1$. Otherwise, $F_0(\mathbf{t}) - F(\mathbf{t}) = d > 0$ as $F_0(\mathbf{t}-) = 1$. Since $F_0(\mathbf{s}) < 1$, if we let $\mathbf{x} = \inf\{\delta\mathbf{t} + (1-\delta)\mathbf{s} : F_0(\mathbf{t}) = F_0(\delta\mathbf{t} + (1-\delta)\mathbf{s}), \delta > 0\}$ then $\mathbf{x}$ is either $\mathbf{t}$ or a member of $(\mathbf{s}, \mathbf{t})$. Also, $\mathbf{x} \in \mathcal{I}_{F_0}$. By continuity of F_0, if $\mathbf{x} = \mathbf{t}$ then for some $\delta > 0$, $\delta\mathbf{t} + (1-\delta)\mathbf{s} \in (\mathbf{s}, \mathbf{t})$ and $0 < F_0(\mathbf{t}) - F_0(\delta\mathbf{t} + (1-\delta)\mathbf{s}) < \frac{d}{2}$, which leads to a contradiction that $\mathcal{D}_1$ contains $\mathcal{Q}_{-\delta}(\mathbf{t})$ with positive μ-measure; otherwise, if $\mathbf{x} \in (\mathbf{s}, \mathbf{t})$, there also exits $\delta > 0$ such that $|F_0(\mathbf{y}) - F_0(\mathbf{t})| < \frac{d}{2}$ for all $\mathbf{y} \in (\mathbf{x} - \delta\mathbf{1}, \mathbf{x}) \subset (\mathbf{s}, \mathbf{t})$, and thus $\mathcal{G}_\delta(\mathbf{x})$ with positive μ-measure is in $\mathcal{D}_1$, a contradiction.

(2) $\partial[\mathbf{s}, \mathbf{t}] \cap \mathcal{D}_1 = \emptyset$. Otherwise, let $\mathbf{x}_0 \in \mathcal{D}_1 \cap \partial[\mathbf{s}, \mathbf{t}]$.

Suppose $\mathbf{x}_0 \in \mathcal{D}_1 \cap \partial_u[\mathbf{s}, \mathbf{t}]$. If $F(\mathbf{x}_0) > F_0(\mathbf{x}_0)$, let $\mathbf{x} = \sup\{\mathbf{x}_0 + \binom{\delta}{0} : F_0(\mathbf{x}_0 + \binom{\delta}{0}) = F_0(\mathbf{x}_0), \delta > 0\}$, then the continuity of F_0 implies that $\mathbf{x} \in \partial_u[\mathbf{s}, \mathbf{t}] \setminus \{\mathbf{t}\}$ and $F_0(\mathbf{x}) < 1$. This fact combining with Condition (d) yields that there exists a $\delta > 0$ such that $-\!\!\bullet\!\!-_\delta(\mathbf{x})$ has positive μ-measure and is a subset of $\mathcal{D}_1$, a contradiction.

Now assume $F_0(\mathbf{x}_0) > F(\mathbf{x}_0)$. Let $\mathbf{x} = \inf\{\mathbf{x}_0 - \binom{\delta}{0} : F_0(\mathbf{x}_0 - \binom{\delta}{0}) = F_0(\mathbf{x}_0), \delta > 0\}$. Then either $\mathbf{x} = \binom{s_1}{t_2}$ or $\mathbf{x} \in \partial_u[\mathbf{s}, \mathbf{t}] \setminus \{\binom{s_1}{t_2}\}$. In the first case, there exists a $\delta > 0$ such that $0 < F_0(\mathbf{x}) - F_0(\mathbf{x} - \binom{0}{\delta}) < \frac{d}{2}$ and $\dagger_{-\delta}(\mathbf{x})$ is contained in $\partial_l[\mathbf{s}, \mathbf{t}] \setminus \{\mathbf{s}\}$ since $F_0(\mathbf{x}_0) > F(\mathbf{x}_0) \geq F(\mathbf{s}) = F_0(\mathbf{s})$. Then $\dagger_{-\delta}(\mathbf{x})$ has positive μ-measure and is contained in $\mathcal{D}_1$. In the second case, there exists a subset of $\mathcal{D}_1$ with positive μ-measure, namely, $\mathrel{-\!\!\bullet\!\!-}_{-\delta}(\mathbf{x})$ for some $\delta > 0$, a contradiction.

Similarly, if $\mathbf{x}_0 \in \mathcal{D}_1$ is contained in some other boundary of $[\mathbf{s}, \mathbf{t}]$, we can show the same contradiction. Hence $\partial[\mathbf{s}, \mathbf{t}] \setminus \{\mathbf{t}\}$ is not in $\mathcal{D}_1$.

(3) In view of the first part in the proof of Proposition 3.8, $\mathcal{D}_1 \cap (\mathbf{s}, \mathbf{t}) = \emptyset$, otherwise, we can find $\mathbf{x} \in \mathcal{I}_{F_0}$ such that $\mathbf{x} \in (\mathbf{s}, \mathbf{t})$ and construct an open square around it with positive μ-measure that is also contained in $\mathcal{D}_1$.

For other cases arisen from Condition (a) and (b), similar argument as (1) - (3) above will lead to a contradiction if $\mathcal{D}_1 \neq \emptyset$.

Now, we have shown that F_n converges pointwisely to F_0 in $[\mathbf{s}, \mathbf{t}]$. By assumption, F_0 is continuous on the bounded close set $[\mathbf{s}, \mathbf{t}]$. Let $\epsilon > 0$. Similar to the second part in the proof of Proposition 3.8, we can select finitely many quantiles $\{a_0, a_1, ..., a_\alpha\}$ such that $a_0 = s_1$, $a_\alpha = t_1$ and $F_0(\binom{a_i}{t_2}) - F_0(\binom{a_{i-1}}{t_2}) < \epsilon$ for all $i = 1, ..., \alpha$, and quantiles $\{b_0, b_1, ..., b_\beta\}$ such that $b_0 = s_2$, $b_\beta = t_2$ and $F_0(\binom{t_1}{b_j}) - F_0(\binom{t_1}{b_{j-1}}) < \epsilon$ for all $j = 1, ..., \beta$. Then there exist $N > 0$ satisfying $|F_n(\binom{a_i}{b_j}) - F_0(\binom{a_i}{b_j})| < \epsilon$ for all $n > N$, and for all $i = 0, ..., \alpha$ and $j = 0, ..., \beta$. For each $\mathbf{x}_0 \in [\mathbf{s}, \mathbf{t}]$, $\mathbf{x}_0 \in [\binom{a_i}{b_j}, \binom{a_{i+1}}{b_{j+1}}]$ for some i, j. Thus, $|F_n(\mathbf{x}_0) - F_m(\mathbf{x}_0)| \leq 12\epsilon$ for all $n, m > N$ by an argument similar to (4.1). Since ϵ is arbitrary, we obtain the uniform convergence in the closed rectangle $[\mathbf{s}, \mathbf{t}]$. ∎

References

Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W. T. and Silverman, E. (1955). An empirical distribution function for sampling incomplete information. *Ann. Math. Statist.*, 641-647.

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, NY.

Groeneboom, P. and Wellner, J.A. (1992). Information bounds and nonparametric maximum likelihood estimation. *Birkhäuser Verlag, Basel.*

The Italian-American Cataract Study Group (1994). Incidence and progression of cortical, nuclear, and posterior subcapsular cataracts. *Amer. J. Ophthal.*, 623-631.

Melbye, M., Biggar, R. J., Ebbesen, P., Sarngadharan, M. G., Weiss, S. H., Gallo, R. C. and Blattner, W. A. (1984). Seroepidemiology of HTLV-III antibody in Danish homosexual men: prevalence, transmission, and disease outcome. *British Med. J.*, **289**, 573-575.

Rudin, W. (1976). Principles of Mathematical Analysis. McGraw-Hill. NY.

Schick, A. and Yu, Q. Q. (2000) Consistency of the *GMLE* with mixed case interval-censored data. *Scan. J. of Statist.* 45-55.

van der Vaart, A. and Wellner, J.A. (2000). Preservation theorems for Glivenko-Cantelli and Uniform Glimvenko-Cantelli Class. *Hign Dimensional probablity II.* E. Gine, D.M. Mason and J.A. Wellner, editors. *Birkhäuser Boston.* 115-133.

Wellner, J.A. (1995). Interval censoring case 2: alternative hypotheses. In *Analysis of censored data*, Proceedings of the workshop on analysis of censored data, Dec. 28, 1994-Jan. 1, 1995, University of Pune, Pune, India. *IMS Lecture Notes, Monograph Series.* **27**, 271-291. H. L. Koul and J. V. Deshpande, editors.

Wellner, J.A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data *Ann Stat.* 779-814.

Wong, G. Y. C. and Yu, Q. Q. (1999). Generalized MLE Of a joint distribution function with multivariate interval-censored data. *J. of Multi. Anal.* 155-166.

Yu, S.H. (2000). Consistency of the generalized mle with multivariate mixed case interval-censored data. *Ph.D dissertation.* Binghamton University.

Yu, Q. Q., Schick, A., Li, L. X. and Wong, G.Y.C. (1998). Asymptotic properties of the GMLE with case 2 interval-censored data. *Statist. & Prob. Let,* 223-228.

5. Appendix: Proof of Theorem 2.1

A proof of Theorem 2.1 when $d = 2$ is given in Example 1 of van der Vaart and Wellner (2000). For readers who are difficult to find their paper, we present a different proof of the theorem, which is in Yu (2000). Let $\mathcal{L}(F) = E(\log \mu_F((L_1, R_1] \times (L_2, R_2]))$. Notice that $\mathcal{L}_n(F) \to \mathcal{L}(F)$ almost surely as $n \to \infty$ by the strong law of large numbers (*SLLN*). We can further write

$$\mathcal{L}(F) = \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} P\{\mathbf{K} = \mathbf{k}\} \cdot E\big(h_{F,\mathbf{k}}(\mathbf{Y}) \mid \mathbf{K} = \mathbf{k}\big), \text{ where}$$

$$\mathbf{k} = \binom{k_1}{k_2} \text{ and } h_{F,\mathbf{k}}(\mathbf{y}) = \sum_{i=0}^{k_1} \sum_{j=0}^{k_2} \mu_{F_0}((y_{1,k_1,i}, y_{1,k_1,i+1}] \times (y_{2,k_2,j}, y_{2,k_2,j+1}]) \cdot$$

$$\cdot \log \mu_F\big((y_{1,k_1,i}, y_{1,k_1,i+1}] \times (y_{2,k_2,j}, y_{2,k_2,j+1}]\big),$$

$$\text{for } -\infty = y_{\delta,k_\delta,0} < y_{\delta,k_\delta,1} < \ldots < y_{\delta,k_\delta,k_\delta} < y_{\delta,k_\delta,k_\delta+1} = \infty, \delta = 1, 2.$$

Define $0 \log 0 = 0$ and $\log 0 = -\infty$ so that the above make sense. Since $\sum_i \sum_j s_{ij} \ln t_{ij}$ is a concave function in the probability vector $(t_{11}, \ldots,)$ where $\sum_{ij} t_{ij} = 1$ and $(s_{11}, \ldots)$ is also a probability vector,

$$h_{F,\mathbf{k}}(\mathbf{y}) \text{ has a maximizer } F \in \mathcal{F} \text{ if and only if } F \text{ satisfies that} \tag{5.1}$$

$$\mu_F((y_{1,k_1,i}, y_{1,k_1,i+1}] \times (y_{2,k_2,j}, y_{2,k_2,j+1}]) = \mu_{F_0}((y_{1,k_1,i}, y_{1,k_1,i+1}] \times (y_{2,k_2,j}, y_{2,k_2,j+1}]),$$

where $i = 0, \ldots, k_1$ and $j = 0, \ldots, k_2$, for each array of real numbers $y_{\delta,k_\delta,1} < \ldots < y_{\delta,k_\delta,k_\delta}$, and for each vector of positive integers $\mathbf{k}$. Also, $|\mathcal{L}(F_0)|$ is bounded by $4E(K_1K_2)$, since $|h_{F_0,\mathbf{k}}| < (k_1+1)(k_2+1)$ in light of the fact that $sup\{|a \log a| : 0 \le a \le 1\} < 1$. This implies that $|\mathcal{L}(F_0)|$ is finite. Thus,

$$F_0 \text{ maximizes } \mathcal{L}(\cdot) \text{ over the set } \mathcal{F}, \text{ and}$$

$$\text{if } F \in \mathcal{F} \text{ maximizes } \mathcal{L}(\cdot) \text{ then } \int_{\mathbb{R}^2} |F(\mathbf{x}) - F_0(\mathbf{x})| \; d\mu(\mathbf{x}) = 0. \tag{5.2}$$

The second statement of (5.2) follows from (5.1).

Let $\mathcal{T}^4 = \{(l_1, r_1, l_2, r_2)' : \binom{-\infty}{-\infty} \le \binom{l_1}{l_2} \le \binom{r_1}{r_2} \le \binom{\infty}{\infty}\}$. We will construct a countable collection $\mathcal{U}$ of Borel subsets of $\mathcal{T}^4$. Let ρ be an arbitrary positive integer. We first select marginal "quantiles" such that

$$-\infty = a_0 < \ldots < a_{\gamma_{\rho,1}} = \infty \; ,$$

$$\text{and for each } i = 1, \ldots, \gamma_{\rho,1}, \mu((-\infty, x] \times [-\infty, \infty]) < i2^{-\rho-1} \quad \text{for all } x < a_i \; ,$$

$$\text{and } \mu((-\infty, y] \times [-\infty, \infty]) \ge i2^{-\rho-1} \quad \text{for all } y > a_i.$$

The selection of a_i's guarantees that $\mu((a_{i-1}, a_i) \times [-\infty, \infty]) < 2^{-\rho-1}$. Similarly, take $-\infty = b_0 < \ldots < b_{\gamma_{\rho,2}} = \infty$ so that $\mu([-\infty,\infty] \times (b_{j-1}, b_j)) < 2^{-\rho-1}$ for each $j = 1, \ldots, \gamma_{\rho,2}$. Since $\mu(\mathbb{R}^2) \leq E(K_1K_2) < \infty$, $\gamma_{\rho,1}$ and $\gamma_{\rho,2}$ are both finite. In fact, $\gamma_{\rho,j} \leq \mu(\mathbb{R}^2)2^{\rho+1}$, $j = 1, 2$. Then we let $\mathcal{Y}_\rho = \{q_{\rho,0}, \ldots, q_{\rho,\beta_\rho}\}$ be the ordered set of all distinct values of $\{a_0, \ldots, a_{\gamma_{\rho,1}}, b_0, \ldots, b_{\gamma_{\rho,2}}\}$. Verify that $\mu((q_{\rho,i-1}, q_{\rho,i}) \times [-\infty,\infty]) < 2^{-\rho-1}$ and $\mu([-\infty,\infty] \times (q_{\rho,j-1}, q_{\rho,j})) < 2^{-\rho-1}$, where $i, j = 1, \ldots, \beta_\rho$. For convenience, let

$$S_{\rho,i,j} := \left(q_{\rho,i-1}, q_{\rho,i}\right] \times \left(q_{\rho,j-1}, q_{\rho,j}\right] \setminus \left\{ \begin{pmatrix} q_{\rho,i} \\ q_{\rho,j} \end{pmatrix} \right\}, \text{ where } i, j = 1, \ldots, \beta_\rho.$$

Notice $\mu(S_{\rho,i,j}) \leq \mu((q_{\rho,i-1}, q_{\rho,i}) \times (-\infty,\infty]) + \mu((-\infty,\infty] \times (q_{\rho,j-1}, q_{\rho,j})) < 2^{-\rho}$. Define sets $V_{\rho,0}, \ldots, V_{\rho,2\beta_\rho}$ such that $V_{\rho,2i-1}$ is the open interval $(q_{\rho,i-1}, q_{\rho,i})$ for each $i = 1, \ldots, \beta_\rho$, and $V_{\rho,2j}$ is $[q_{\rho,j}, q_{\rho,j}]$ for each $j = 0, \ldots, \beta_\rho$. Let $W_{\rho,i,j} = V_{\rho,i} \times V_{\rho,j}$, for $i, j = 0, \ldots, 2\beta_\rho$. Finally, let $\mathcal{U} = \cup_\rho \mathcal{U}_\rho$, where $\mathcal{U}_\rho$ is the collection of all nonempty sets of the form $U_{\rho,\mathbf{i},\mathbf{j}} = (W_{\rho,i_1,j_1} \times W_{\rho,i_2,j_2}) \cap \mathcal{T}^4$, for $0 \leq i_1 \leq j_1 \leq 2\beta_\rho$ and $0 \leq i_2 \leq j_2 \leq 2\beta_\rho$.

Let $\mathbb{Q}^*$ be the union of $\mathcal{A}^*$ and $\mathbb{Q}^2$, where $\mathcal{A}^* = \cup_{\rho \in \mathbb{Z}^+} {\mathcal{Y}_\rho}^2$ and $\mathbb{Q}^2$ is the set of all points in $\mathbb{R}^2$ whose coordinates are rational. Then for each $\omega \in \Omega$, there exists a subsequence $\{n'\}$ of $\{n\}$ tending to infinity such that $\hat{F}_{n'}(\mathbf{x};\omega) \to F(\mathbf{x};\omega)$ for all $\mathbf{x} \in \mathbb{Q}^*$, where $F \in \mathcal{F}$. To uniquely determine the F, for each $\mathbf{x} \in \mathbb{R}^2 \setminus \mathbb{Q}^*$, define $F_\omega(\mathbf{x}) = F(\mathbf{x};\omega) = \inf\{F(\mathbf{a};\omega) : \mathbf{a} \in \mathbb{Q}^* \text{ and } \mathbf{x} \leq \mathbf{a}\}$. Since $\hat{F}_n(\cdot;\omega)$ is a distribution function for each n and each ω, F_ω is nondecreasing in each variable and bounded by 0 and 1, obviously. Furthermore, $\mu_{F_\omega}(W)$ has nonnegative value for each $W \in \mathcal{W}$ (an algebra of $\mathbb{R}^2$ defined in Section 2), and thus, μ_{F_ω} is a measure induced by F_ω and F_ω is a member of $\mathcal{F}$. Fix an $\omega \in \Omega$. Let $\epsilon > 0$. Note that if $\mathbf{x}$ is a continuity point of F_ω, then there exist $\mathbf{q}_1, \mathbf{q}_2 \in \mathbb{Q}^2$ such that $\mathbf{q}_1 < \mathbf{x} < \mathbf{q}_2$ and $F(\mathbf{x};\omega) - \epsilon < F(\mathbf{q}_1;\omega) \leq F(\mathbf{q}_2;\omega) < F(\mathbf{x};\omega) + \epsilon$. For each n, we have $\hat{F}_n(\mathbf{q}_1;\omega) \leq \hat{F}_n(\mathbf{x};\omega) \leq \hat{F}_n(\mathbf{q}_2;\omega)$. Then, $F(\mathbf{x};\omega) - \epsilon \leq \liminf_{n'} \hat{F}_{n'}(\mathbf{x};\omega) \leq \limsup_{n'} \hat{F}_{n'}(\mathbf{x};\omega) \leq F(\mathbf{x};\omega) + \epsilon$. Since ϵ is arbitrary, $F(\mathbf{x};\omega) = \lim_{n'} \hat{F}_{n'}(\mathbf{x};\omega)$. Let

$$\mathcal{D}_\omega = \{\mathbf{x} \in \mathbb{R}^2 : |\limsup_{n'} \hat{F}_{n'}(\mathbf{x};\omega) - F(\mathbf{x};\omega)| \vee |\liminf_{n'} \hat{F}_{n'}(\mathbf{x};\omega) - F(\mathbf{x};\omega)| > 0\},$$

then $\mathcal{D}_\omega$ does not contain any continuity point of F_ω. If $\mathbf{x} \in \mathcal{D}_\omega$, verify that one of the following must be true:

1. $\mu_{F_\omega}(\{\mathbf{x}\}) > 0$,
2. $\mathbf{x}$ is a horizontal continuity points of F_ω (as defined in Section 3),
3. $\mathbf{x}$ is a vertical continuity point of F_ω.

Let $\mathcal{D}_{i,\omega}$ be the set of all points in $\mathcal{D}_\omega$ satisfying the ith condition above. Then $\mathcal{D}_\omega = \mathcal{D}_{1,\omega} \cup \mathcal{D}_{2,\omega} \cup \mathcal{D}_{3,\omega}$. We shall show next that $\mu(\mathcal{D}_\omega) = 0$.

1. Suppose $\mathcal{D}_{1,\omega} \neq \emptyset$. If $\mathbf{x} \in \mathcal{D}_{1,\omega}$, then $\mathbf{x} \notin \mathcal{A}^*$ by the definition of F_ω. For each positive integer ρ, $\mathbf{x} \in S_{\rho,i_\rho,j_\rho}$ for some i_ρ and j_ρ, and

$$\mu(\{\mathbf{x}\}) \leq \mu(S_{\rho,i_\rho,j_\rho}) < 2^{-\rho} \overset{\rho\to\infty}{\longrightarrow} 0.$$

Since μ_{F_ω} is a finite measure, there are at most countably many elements in $\mathcal{D}_{1,\omega}$, so $\mu(\mathcal{D}_{1,\omega}) = 0$.

2. Suppose $\mathcal{D}_{2,\omega} \neq \emptyset$. By monotonicity of F_ω in each variable, there exists $(\mathbf{a}, \mathbf{b}) \subset \mathcal{D}_{2,\omega}$ such that the second coordinates of $\mathbf{a}$ and $\mathbf{b}$ are the same, $\mu_{F_\omega}((\mathbf{a}, \mathbf{b})) > 0$ and $(\mathbf{a}, \mathbf{b}) \cap \mathcal{A}^* = \emptyset$. Note that $(\mathbf{a}, \mathbf{b})$ is fixed here. For each ρ, $(\mathbf{a}, \mathbf{b})$ is contained either in $(q_{\rho,i_\rho-1}, q_{\rho,i_\rho}) \times [-\infty, \infty]$ or in $[-\infty, \infty] \times (q_{\rho,j_\rho-1}, q_{\rho,j_\rho})$ for some i_ρ or j_ρ. Either way, as $\rho \to \infty$,

$$\begin{aligned}\mu_{F_\omega}((\mathbf{a}, \mathbf{b})) &\leq \mu((q_{\rho,i_\rho-1}, q_{\rho,i_\rho}) \times [-\infty, \infty]) + \mu([-\infty, \infty] \times (q_{\rho,j_\rho-1}, q_{\rho,j_\rho})) \\ &< 2^{-\rho} \longrightarrow 0.\end{aligned}$$

The above implies that $\mu(\mathcal{D}_{2,\omega}) = 0$ as there are at most countably many such $(\mathbf{a}, \mathbf{b})$'s in $\mathcal{D}_{1,\omega}$ by the boundedness of the measure μ_{F_ω}.

3. Suppose $\mathcal{D}_{3,\omega} \neq \emptyset$. By symmetry of $\mathcal{D}_{2,\omega}$, $\mu(\mathcal{D}_{3,\omega}) = 0$.

The above implies that $\lim_{n'} \hat{F}_{n'}(\mathbf{x}; \omega) = F(\mathbf{x}; \omega)$ except on a set with μ-measure 0. Hence, for each $\omega \in \Omega$,

$$\lim_{n' \to \infty} \int_{\mathbb{R}^2} |\hat{F}_{n'}(\mathbf{x}; \omega) - F(\mathbf{x}; \omega)| d\mu(\mathbf{x}) = 0. \tag{5.3}$$

Let $\hat{Q}_n$ denote the empirical estimator of Q, the distribution of $(\mathbf{L}, \mathbf{R})$. By $SLLN$, $\Omega_U = \{\omega : \hat{Q}_n(U; \omega) \to Q(U)\}$ has probability 1 for each Borel subset U of $\mathcal{T}^4$, so does $\Omega' = \{\omega : \mathcal{L}_n(F_0; \omega) \to \mathcal{L}(F_0)\}$. Hence, $\Omega^* = \Omega' \cap (\cap_{U \in \mathcal{U}} \Omega_U)$ has probability 1. Let ω^* be a member of Ω^*. To simplify the notations in this proof, let F denote the function defined by $F(\mathbf{x}) = F(\mathbf{x}; \omega^*)$, F_n the $GMLE$ of the distribution function defined by $F_n(\mathbf{x}) = \hat{F}_n(\mathbf{x}; \omega^*)$, where $\mathbf{x} \in \mathbb{R}^2$, and Q_n the empirical distribution function defined by $Q_n(U) = \hat{Q}_n(U; \omega^*)$, where U is a member of the Borel σ-field $\mathcal{B}(\mathcal{T}^4)$. Without loss of generality ($WLOG$), assume $\{n'\} = \{n\}$. Obviously $\mathcal{L}(F) \leq \mathcal{L}(F_0)$. Also, $\mathcal{L}(F_0) \leq \liminf_{n \to \infty} \mathcal{L}_n(F_n; \omega^*)$, because $\mathcal{L}_n(F_0; \omega^*) \leq \mathcal{L}_n(F_n; \omega^*)$ by the definition of the $GMLE$, and the fact that $\mathcal{L}_n(F_0; \omega^*) \to \mathcal{L}(F_0)$ by the choice of ω^*. If we can show that

$$\limsup_{n \to \infty} \mathcal{L}_n(F_n; \omega^*) \leq \mathcal{L}(F) \tag{5.4}$$

then $\mathcal{L}(F_0) \leq \mathcal{L}(F)$. This will further conclude that $\Omega^0 = \{\omega : \mathcal{L}(F_0) = \mathcal{L}(F(\cdot; \omega))\}$ contains Ω^* by the arbitrary choice of ω^*, and thus has probability 1. In addition,

$$\limsup_{n \to \infty} \int_{\mathbb{R}^2} |\hat{F}_{n'}(\mathbf{x}; \omega) - F_0(\mathbf{x})| \ d\mu(\mathbf{x}) = 0, \text{ for each } \omega \in \Omega^0,$$

in view of (5.2) and (5.3), thus the theorem is proved. Notice that $\mathcal{L}_n(F_n; \omega^*) = \int_{\mathcal{T}^4} \log \ \mu_{F_n}((\mathbf{l}, \mathbf{r}]) dQ_n((\mathbf{l}, \mathbf{r}))$, where $(\mathbf{l}, \mathbf{r})$ denotes the vector $(l_1, r_1, l_2, r_2)'$. The needed inequality (5.4) can be written as

$$\limsup_{n \to \infty} \int_{\mathcal{T}^4} \log \mu_{F_n}((\mathbf{l}, \mathbf{r}]) dQ_n((\mathbf{l}, \mathbf{r})) \leq \int_{\mathcal{T}^4} \log \mu_F((\mathbf{l}, \mathbf{r}]) dQ((\mathbf{l}, \mathbf{r})). \tag{5.5}$$

We now show that (5.5) holds for each $\omega^* \in \Omega^*$. Fix a positive integer ρ and a negative integer ϱ. Remember that every element in $\mathcal{U}_\rho$ can be written as $U_{\rho, \binom{i_1}{i_2}, \binom{j_1}{j_2}}$

or $V_{\rho,i_1} \times V_{\rho,j_1} \times V_{\rho,i_2} \times V_{\rho,j_2}$ for some $\binom{i_1}{i_2} \leq \binom{j_1}{j_2}$. Then the following is immediate.

$$\begin{aligned}\int_{\mathcal{T}^4} \log \ \mu_{F_n}(\big(\mathbf{l},\mathbf{r}\big]) \ dQ_n((\mathbf{l},\mathbf{r})) &\leq \int_{\mathcal{T}^4} \varrho \vee \log \ \mu_{F_n}(\big(\mathbf{l},\mathbf{r}\big]) \ dQ_n((\mathbf{l},\mathbf{r})) \\ &\leq \sum_{U \in \mathcal{U}_\rho} M_n(U) \ Q_n(U),\end{aligned}$$

where $M_n(U) = \sup\{\varrho \vee \log \mu_{F_n}(\big(\mathbf{l},\mathbf{r}\big]) : (\mathbf{l},\mathbf{r}) \in \bar{U}\}$ and $\bar{U}$ is the closure of U. For any U in $\mathcal{U}_\rho$, let

$$\begin{aligned}&r_{U,i}{}^+ = \sup\{r_i : (\mathbf{l},\mathbf{r}) \in U\}, \ \ r_{U,i}{}^- = \inf\{r_i : (\mathbf{l},\mathbf{r}) \in U\}, \\ &l_{U,i}{}^+ = \sup\{l_i : (\mathbf{l},\mathbf{r}) \in U\} \ \text{ and } l_{U,i}{}^- = \inf\{l_i : (\mathbf{l},\mathbf{r}) \in U\}, \ \text{ where } i = 1, 2.\end{aligned}$$

Let $\big(\mathbf{l},\mathbf{r}\big]_U{}^+ = (l_{U,1}{}^-, r_{U,1}{}^+] \times (l_{U,2}{}^-, r_{U,2}{}^+]$ and $\big(\mathbf{l},\mathbf{r}\big]_U{}^- = (l_{U,1}{}^+, r_{U,1}{}^-] \times (l_{U,2}{}^+, r_{U,2}{}^-]$ for convenience. It can be shown that $M_n(U) \leq \varrho \vee \log \mu_{F_n}(\big(\mathbf{l},\mathbf{r}\big]_U{}^+)$. Thus,

$$M_n(U) \to M(U) = \varrho \vee \sup_{(\mathbf{l},\mathbf{r}) \in \bar{U}} \log \mu_F(\big(\mathbf{l},\mathbf{r}\big]) \leq \varrho \vee \log \mu_F(\big(\mathbf{l},\mathbf{r}\big]_U{}^+). \tag{5.6}$$

By the choice of ω^*, $Q_n(U) \to Q(U)$. Hence, $\sum_{U \in \mathcal{U}_\rho} M_n(U) Q_n(U) \to \sum_{U \in \mathcal{U}_\rho} M(U) Q(U)$. Let $m_n(U) = \inf\{\varrho \vee \log \mu_{F_n}(\big(\mathbf{l},\mathbf{r}\big]) : (\mathbf{l},\mathbf{r}) \in \bar{U}\}$. Similarly, we obtain the following:

$$m_n(U) \to m(U) = \varrho \vee \inf_{(\mathbf{l},\mathbf{r}) \in \bar{U}} \log \mu_F(\big(\mathbf{l},\mathbf{r}\big]) \geq \varrho \vee \log \mu_F(\big(\mathbf{l},\mathbf{r}\big]_U{}^-) \tag{5.7}$$

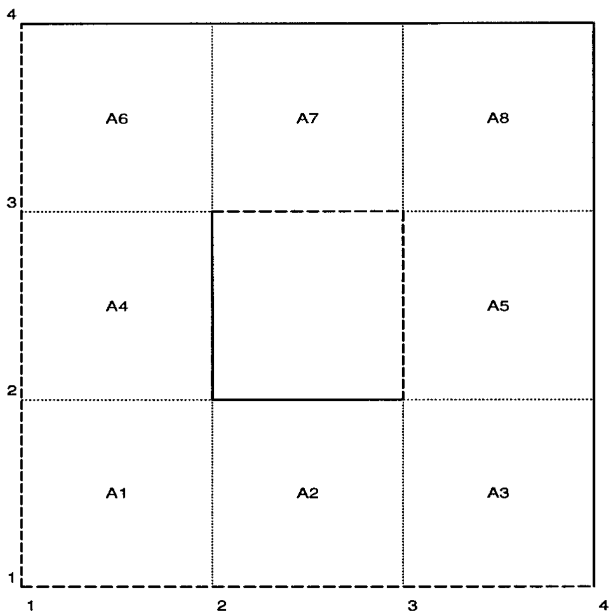
and $\sum_{U \in \mathcal{U}_\rho} m_n(U) Q_n(U) \to \sum_{U \in \mathcal{U}_\rho} m(U) Q(U)$. Verify

$$\begin{aligned}M(U) - m(U) &\leq \varrho \vee \log \mu_F(\big(\mathbf{l},\mathbf{r}\big]_U{}^+) - \varrho \vee \log \mu_F(\big(\mathbf{l},\mathbf{r}\big]_U{}^-) \ \text{(by (5.6) and (5.7))} \\ &\leq e^{-\varrho} \cdot [\mu_F(\big(\mathbf{l},\mathbf{r}\big]_U{}^+) - \mu_F(\big(\mathbf{l},\mathbf{r}\big]_U{}^-)], \ \text{where } U \in \mathcal{U}_\rho.\end{aligned}$$

If L_1 or L_2 belongs to $(q_{\rho,i-1}, q_{\rho,i})$ for some $i = 1, ..., \beta_\rho$, then $\mu_F(\big(\mathbf{l},\mathbf{r}\big]_U{}^+) - \mu_F(\big(\mathbf{l},\mathbf{r}\big]_U{}^-)$ can be expressed as the sum of at most $3^2 - 1$ μ-measures of the rectangles whose μ-measures are no more than the μ-measures of the rectangles of the form $(q_{\rho,i-1}, q_{\rho,i}) \times (-\infty, \infty]$ or $(-\infty, \infty] \times (q_{\rho,i-1}, q_{\rho,i})$, where $i = 1, ..., \beta_\rho$. For instance, given

FIGURE 1. $\mu_F(\big(\mathbf{1},\mathbf{41}\big]) - \mu_F(\big(\mathbf{21},\mathbf{31}\big]) = \sum_{i=1}^{8} \mu_F(A_i)$

Note: *For convenience, let* $q_{\rho,i_1-1} = q_{\rho,i_2-1} = 1$, $q_{\rho,i_1} = q_{\rho,i_2} = 2$, $q_{\rho,j_1-1} = q_{\rho,j_2-1} = 3$ *and* $q_{\rho,j_1} = q_{\rho,j_2} = 4$.



$$
\begin{aligned}
&U = (q_{\rho,i_1-1}, q_{\rho,i_1}) \times (q_{\rho,i_2-1}, q_{\rho,i_2}) \times (q_{\rho,j_1-1}, q_{\rho,j_1}) \times (q_{\rho,j_2-1}, q_{\rho,j_2}),\\
&\mu_F(\big(\mathbf{l},\mathbf{r}\big]_U{}^{+}) - \mu_F(\big(\mathbf{l},\mathbf{r}\big]_U{}^{-})\\
= &\left[F\Big(\begin{pmatrix} q_{\rho,j_1} \\ q_{\rho,j_2} \end{pmatrix}\Big) + F\Big(\begin{pmatrix} q_{\rho,i_1-1} \\ q_{\rho,i_2-1} \end{pmatrix}\Big) - F\Big(\begin{pmatrix} q_{\rho,j_1} \\ q_{\rho,i_2-1} \end{pmatrix}\Big) - F\Big(\begin{pmatrix} q_{\rho,i_1-1} \\ q_{\rho,j_2} \end{pmatrix}\Big)\right]\\
&- \left[F\Big(\begin{pmatrix} q_{\rho,j_1-1} \\ q_{\rho,j_2-1} \end{pmatrix}\Big) + F\Big(\begin{pmatrix} q_{\rho,i_1} \\ q_{\rho,i_2} \end{pmatrix}\Big) - F\Big(\begin{pmatrix} q_{\rho,j_1-1} \\ q_{\rho,i_2} \end{pmatrix}\Big) - F\Big(\begin{pmatrix} q_{\rho,i_1} \\ q_{\rho,j_2-1} \end{pmatrix}\Big)\right]\\
= &\sum_{i=1}^{8} \mu_F(A_i) \text{ as illustrated in Figure 1.}
\end{aligned}
$$

Thus if L_1 or L_2 belongs to $(q_{\rho,i-1}, q_{\rho,i})$ for some $i = 1, ..., \beta_\rho$, then $M(U) - m(U) > \frac{8}{\rho}$ implies that at least one of the rectangles $S_{\rho,i,j}$, $j = 1, ..., \beta_\rho$, has μ_F-measure exceeding e^ϱ/ρ. Similarly, if R_1 or R_2 is in an interval $(q_{\rho,j-1}, q_{\rho,j})$ for some $j = 1, ..., \beta_\rho$, then the same implication is true for at least one of the rectangles $S_{\rho,i,j}$, $i = 1, ..., \beta_\rho$. If U

contains only one point, then $M(U)-m(U)=0$. It follows that

$$
\begin{aligned}
&\sum_{U\in\mathcal{U}_\rho}(M(U)-m(U))Q(U)\\
&\le \frac{8}{\rho}\sum_{U\in\mathcal{U}_\rho}Q(U)I\{M(U)-m(U)\le\frac{8}{\rho}\}+|\varrho|\sum_{U\in\mathcal{U}_\rho}Q(U)I\{M(U)-m(U)>\frac{8}{\rho}\}\\
&\le \frac{8}{\rho}+|\varrho|\sum_{i=1}^{\beta_\rho}P\{q_{\rho,i-1}<L_1<q_{\rho,i}\}\sum_{j=1}^{\beta_\rho}I\{\mu_F(S_{\rho,i,j})>e^{\varrho}/\rho\}\\
&\quad+|\varrho|\sum_{i=1}^{\beta_\rho}P\{q_{\rho,i-1}<L_2<q_{\rho,i}\}\sum_{j=1}^{\beta_\rho}I\{\mu_F(S_{\rho,i,j})>e^{\varrho}/\rho\}\\
&\quad+|\varrho|\sum_{j=1}^{\beta_\rho}P\{q_{\rho,j-1}<R_1<q_{\rho,j}\}\sum_{i=1}^{\beta_\rho}I\{\mu_F(S_{\rho,i,j})>e^{\varrho}/\rho\}\\
&\quad+|\varrho|\sum_{j=1}^{\beta_\rho}P\{q_{\rho,j-1}<R_2<q_{\rho,j}\}\sum_{i=1}^{\beta_\rho}I\{\mu_F(S_{\rho,i,j})>e^{\varrho}/\rho\}
\end{aligned}
\tag{5.8}
$$

Notice that $\mu_F(\mathbb{R}^2)$ is bounded by 1, so there are no more than $\rho e^{-\varrho}+1$ terms of the form $\mu_F(S_{\rho,i,j})$, $i,j=1,...,\beta_\rho$, exceeding e^{ϱ}/ρ. Furthermore,

$$
\begin{aligned}
&\sum_{j=1}^{2}[P\{q_{\rho,i-1}<L_j<q_{\rho,i}\}+P\{q_{\rho,i-1}<R_j<q_{\rho,i}\}]\\
&\le 2\mu((q_{\rho,i-1},q_{\rho,i})\times[-\infty,\infty])+2\mu([-\infty,\infty]\times(q_{\rho,i-1},q_{\rho,i}))\\
&\le 4\cdot 2^{-\rho-1}=2^{1-\rho}.
\end{aligned}
$$

Then, (5.8) can be rewritten as

$$
\sum_{U\in\mathcal{U}_\rho}(M(U)-m(U))Q(U)\le\frac{8}{\rho}+|\varrho|2^{1-\rho}(\rho e^{-\varrho}+1).
$$

Therefore, (5.5) follows from the following inequality.

$$
\begin{aligned}
&\limsup_{n\to\infty}\int_{\mathcal{T}^4}\log\mu_{F_n}(\big(\mathbf{l},\mathbf{r}\big])dQ_n((\mathbf{l},\mathbf{r}))\\
&\le\int_{\mathcal{T}^4}\varrho\vee\log\mu_F(\big(\mathbf{l},\mathbf{r}\big])dQ((\mathbf{l},\mathbf{r}))+\frac{8}{\rho}+|\varrho|2^{1-\rho}(\rho e^{-\varrho}+1)\\
&\overset{\rho\to\infty}{\longrightarrow}\int_{\mathcal{T}^4}\varrho\vee\log\mu_F(\big(\mathbf{l},\mathbf{r}\big])dQ((\mathbf{l},\mathbf{r}))\\
&\overset{\varrho\to-\infty}{\longrightarrow}\int_{\mathcal{T}^4}\log\mu_F(\big(\mathbf{l},\mathbf{r}\big])dQ((\mathbf{l},\mathbf{r}))\quad\blacksquare
\end{aligned}
$$